



UNIVERSIDAD CARLOS III DE MADRID

## **TESIS DOCTORAL**

# **Métodos estadísticos en series temporales no lineales, con aplicación a la predicción de energía eólica**

Autor:

Miguel Ángel Bermejo

Directores:

Daniel Peña e Ismael Sánchez

**DEPARTAMENTO DE ESTADÍSTICA**

Leganés, Junio 2011



*a mis padres y hermana*



# Agradecimientos

Durante el inicio de mi vida académica no fui un alumno especialmente aplicado. Por ello, mi primer agradecimiento ha de ser para aquellos profesores que durante esos años creyeron en mí a pesar de mi falta de atención y de esfuerzo y me animaron y ayudaron para continuar. Poco a poco entre todos lograron que fuera convirtiéndome en mejor estudiante.

Me gustaría agradecer especialmente su ayuda a mis directores, Daniel Peña e Ismael Sánchez. La primera vez que se me ocurrió la idea de realizar un doctorado y de dedicarme a la investigación fue durante el curso de series temporales que Daniel impartía en la licenciatura de Estadística, por lo que a él le debo el estar hoy aquí. Además gracias a su sugerencia se incorporó como director Ismael que durante estos años ha sido una gran ayuda y un gran apoyo. En estos cuatro años trabajando junto a ellos he aprendido mucho como estudiante y como persona.

También quiero agradecer al profesor James W. Taylor y a su entonces estudiante Jooyoung Jeon su gran acogida y las diferentes charlas que me ayudaron a progresar en mis inicios en la tesis durante mi estancia en la Universidad de Oxford. Asimismo me gustaría agradecer a los miembros del proyecto Safewind que con sus comentarios en las diferentes reuniones del proyecto han ayudado a la mejora de este trabajo.

Llevo cursando titulaciones del departamento de Estadística muchos años: la diplomatura, la licenciatura, el máster y el doctorado, por lo que creo que puedo decir que he aprendido algo de todos los profesores del departamento. Por ello, siempre le estaré muy agradecido a un departamento al que entré, sin saber muy bien que era eso de Estadística que había elegido tras la selectividad y salgo tras completar un ciclo académico inimaginable para mí el día que asistí a clase por primera vez. Dentro de la gente que he podido conocer en el departamento me gustaría agradecer especialmente su apoyo y amistad (y los partidos de squash) a mis compañeros de doctorado que al fin y al cabo son las únicas personas que consiguen comprenderte durante este periodo. Además de llevarme un título, me llevo muchos amigos, gente como Nuria, Kenedy, Henry, Sofía, Javi, Cristina, Alberto, Leo, Gabi, etc etc etc. Entre todos ellos me gustaría agradecer especialmente a Roberto los grandes y duros momentos que pasamos durante el máster

y el inicio de la tesis y lamentar que hoy no pueda estar en el departamento terminando su tesis como yo.

Para los amigos que me conocen desde hace bastante tiempo, que esté terminando un doctorado es algo impensable. Durante estos años he cambiado mucho, pero ellos siguen ahí. Sin todos los momentos pasados junto a ellos estos años hubiesen sido insoportables. Por ello me gustaría agradecerse a todos, desde los que llevan ahí desde los años del instituto, gente como Rubén (por tres), Jose Ángel, Javi, Jose Luís, Oscar o David, hasta los que fueron apareciendo durante los años posteriores (para ellos no será tan sorprendente que termine) como Laura, Kike o Pedroche.

Para el final quedan mis padres y mi hermana Sara. La palabra agradecimiento se queda corta en este caso, ellos son sin duda lo mejor y lo más importante de mi vida. Si hoy puedo terminar mi doctorado es únicamente posible gracias al esfuerzo y dedicación de mis padres, que trabajaron muy duro para darnos a mi hermana y a mi las posibilidades que ellos no pudieron tener. Les estaré eternamente agradecido ya que gracias a ellos, a su esfuerzo y a su cariño me he podido convertir en la persona que hoy soy.

# Índice

<b>Lista de figuras</b>	<b>vii</b>
<b>Lista de tablas</b>	<b>xi</b>
<b>Lista de acrónimos</b>	<b>xiii</b>
<b>1 Introducción</b>	<b>1</b>
1.1 Modelos no lineales para series temporales . . . . .	2
1.1.1 Modelos por umbrales . . . . .	2
1.1.2 Modelos autorregresivos exponenciales de amplitud dependiente .	3
1.1.3 Modelos bilineales . . . . .	3
1.1.4 Modelos de heterocedasticidad condicional . . . . .	4
1.1.5 Modelos de segunda generación . . . . .	5
1.1.6 Modelos no paramétricos . . . . .	6
1.2 Problemas tratados en la tesis . . . . .	7
1.2.1 Identificación de modelos no lineales . . . . .	7
1.2.2 Predicción con modelos no lineales . . . . .	12
1.2.3 Aplicaciones en energía eólica . . . . .	13
1.3 Estructura de la tesis . . . . .	16
<b>2 Modelos autorregresivos por umbrales</b>	<b>19</b>
2.1 Tipos de modelos por umbrales . . . . .	20
2.2 Estacionariedad en modelos SETAR . . . . .	22
2.3 Estimación de modelos SETAR . . . . .	23
2.3.1 Mínimos cuadrados ordinarios . . . . .	23
2.3.2 Propiedades del estimador OLS . . . . .	25
2.4 Selección de la estructura autorregresiva . . . . .	28
2.4.1 Criterio de información de Akaike . . . . .	28
2.4.2 Criterios de información Bayesianos . . . . .	29

2.4.3	Criterios basados en Validación Cruzada . . . . .	30
2.4.4	Criterios de selección Bootstrap . . . . .	31
2.4.5	Mejora de criterios de selección . . . . .	32
2.5	Identificación de modelos SETAR . . . . .	33
2.5.1	Contrastes de linealidad . . . . .	33
2.5.2	Algoritmos de búsqueda intensiva . . . . .	41
2.6	Predicción con modelos SETAR . . . . .	44
<b>3</b>	<b>Identificación de modelos SETAR mediante estimación recursiva</b>	<b>47</b>
3.1	Modelo autorregresivo ordenado con parámetros variables en el tiempo .	47
3.2	Estimación recursiva de de parámetros variables en el tiempo . . . . .	49
3.2.1	Mínimos cuadrados recursivos y ponderados . . . . .	49
3.2.2	Factores de olvido adaptativos . . . . .	51
3.2.3	Propiedades del estimador RLS con factor de olvido . . . . .	53
3.3	Herramienta de identificación, ARLS . . . . .	54
3.3.1	Estimación inicial de la recursividad . . . . .	55
3.3.2	Justificación del factor de olvido . . . . .	56
3.3.3	Intervalos de confianza de $\phi_{st}$ . . . . .	57
3.3.4	Rendimiento en muestras finitas de los intervalos asintóticos . . .	58
3.3.5	Eficacia del método ARLS en muestras finitas . . . . .	60
3.4	Procedimiento automático, Aut-ARLS . . . . .	64
3.5	Aplicación en datos reales clásicos . . . . .	66
3.5.1	Linces canadienses . . . . .	67
3.5.2	Manchas solares . . . . .	68
3.5.3	Producto nacional bruto EEUU . . . . .	70
3.6	Conclusiones y líneas futuras de investigación . . . . .	72
<b>4</b>	<b>Densidad predictiva basada en momentos condicionados y máxima entropía.</b>	
	<b>Aplicación a la predicción de potencia eólica</b>	<b>73</b>
4.1	Energía eólica . . . . .	73
4.1.1	Análisis descriptivo de la distribución de la potencia generada . .	74
4.1.2	Notación . . . . .	75
4.2	Densidad predictiva en energía eólica . . . . .	76
4.3	Estimación de los momentos condicionados . . . . .	78
4.3.1	Método de estimación recursiva . . . . .	79
4.4	Modelización adaptativa de densidades predictivas . . . . .	80
4.4.1	Distribución Normal truncada . . . . .	80
4.4.2	Distribución Normal censurada . . . . .	83



4.4.3	Distribución Beta . . . . .	85
4.4.4	Distribución de Máxima Entropía . . . . .	87
4.4.5	Comparación de las distribuciones empleadas . . . . .	89
4.5	Criterios de evaluación de densidades predictivas . . . . .	90
4.5.1	Comprobación de la estimación . . . . .	90
4.5.2	Verificación de las predicciones . . . . .	92
4.6	Selección de la distribución óptima empleando criterios de evaluación . . .	93
4.7	Comportamiento de la estimación en diferentes parques eólicos . . . . .	94
4.8	Conclusiones y líneas futuras . . . . .	95
<b>5</b>	<b>Predicción de producción eólica empleando modelos SETAR</b>	<b>105</b>
5.1	Predicciones a corto plazo de potencia eólica . . . . .	105
5.2	Modelos SETAR en potencia eólica . . . . .	107
5.2.1	Patrón común detectado . . . . .	109
5.2.2	Modelos SETAR . . . . .	111
5.2.3	Modelos SETAR con densidad condicionada variable en el tiempo	112
5.3	Predicción empleando modelos SETAR-TVCD . . . . .	113
5.4	Predicciones puntuales de producción eólica . . . . .	115
5.5	Conclusiones . . . . .	117
<b>6</b>	<b>Contribuciones de la tesis</b>	<b>125</b>
	<b>Referencias</b>	<b>129</b>



# Lista de figuras

1.1	Realización de tamaño 1000 a partir del modelo (1.2) para valores $J = 2$ , $a_0^{(1)} = 0$ , $a_0^{(2)} = 3$ , $a_1^{(1)} = .8$ , $a_1^{(2)} = -.8$ , dependiendo el indicador temporal $J_t$ del valor de $Y_{t-1}$ .	3
1.2	Realización de tamaño 200 del modelo (1.3), con $k = 2$ , $\alpha = (2, -1)$ , $\beta = (.5, -.5)$ y $\delta = 1$ .	4
1.3	Realización de tamaño 200 del modelo (1.4), con $p = q = P = Q = 1$ , $\phi = .8$ , $\theta = -.5$ y $\gamma = .5$ .	4
1.4	Ejemplos de modelos con heterocedasticidad condicional	5
1.5	Gráfico de dispersión de $Y_t$ respecto a $Y_{t-1}$ .	8
1.6	Comparación del gráfico de una serie procedente de modelos lineales y no lineales	9
1.7	Diagramas de dispersión directos para datos provenientes de modelos lineales y no lineales	10
1.8	Ajuste lineal y suavizado no paramétrico sobre un proceso generado a partir de un modelo no lineal	10
1.9	Evolución del sistema eléctrico español	14
1.10	Ejemplo de serie de datos de producción eólica	15
2.1	Gráfico de dispersión de $Y_t$ respecto a la variable umbral $Y_{t-1}$	25
3.1	Ejemplo de funcionamiento del procedimiento ARLS	59
3.2	Curva de potencia para $n=150$	62
3.3	Curva de potencia para $n=500$	62
3.4	Histograma de los umbrales estimados para diferentes valores de $\delta$	63
3.5	Histograma de los umbrales estimados empleando Aut-ARLS sobre 1000 réplicas del modelo (3.30)	65
3.6	Tasa de detección para diferentes valores de $\delta_1$ y $\delta_2$ obtenida empleando Aut-ARLS	66
3.7	Porcentaje de acierto de Aut-ARLS	67

3.8	Serie de los lincs canadienses transformada con logaritmos . . . . .	67
3.9	Detección del umbral en los datos de los lincs canadienses empleando Aut-ARLS . . . . .	68
3.10	Serie de las manchas solares . . . . .	69
3.11	Detección del umbral empleando Aut-ARLS para los datos de las manchas solares . . . . .	69
3.12	Tasa de crecimiento del producto nacional bruto de Estados Unidos desde 1947 a 2003. . . . .	71
4.1	Potencia producida por un parque eólico, datos horarios . . . . .	74
4.2	Gráfico de caja para de la potencia medida respecto a diferentes niveles de la predicción puntual . . . . .	75
4.3	Media de la estimación de los momentos para diferentes horizontes . . . . .	79
4.4	Estimación recursiva de los momentos centrados condicionados para $h=1,3,6,12$ . . . . .	81
4.5	Estimación recursiva de los momentos condicionados para $h=1,3,6,12$ . . . . .	82
4.6	Ejemplo de la senda de densidades predictiva empleando la distribución Normal Truncada para diferentes horizontes . . . . .	83
4.7	Ejemplo de la senda de densidades predictiva empleando la distribución Normal Censurada para diferentes horizontes . . . . .	86
4.8	Ejemplo de la senda de densidades predictiva empleando la distribución Beta para diferentes horizontes . . . . .	87
4.9	Ejemplo de la senda de densidades predictiva empleando la distribución de máxima entropía para diferentes horizontes . . . . .	90
4.10	Densidades predictivas para diferentes instantes de tiempo y horizontes . . . . .	91
4.11	Histograma de los valores PIT obtenidos para diferentes densidades predictivas estimadas para un parque eólico a horizonte $h = 1$ . . . . .	97
4.12	Histograma de los valores PIT obtenidos para diferentes densidades predictivas estimadas para un parque eólico a horizonte $h = 6$ . . . . .	98
4.13	Histograma de los valores PIT obtenidos para diferentes densidades predictivas estimadas para un parque eólico a horizonte $h = 12$ . . . . .	99
4.14	Histograma de los valores PIT obtenidos para diferentes densidades predictivas estimadas para un parque eólico a horizonte $h = 24$ . . . . .	100
4.15	Valor del Brier Score obtenido para las diferentes distribuciones estimadas en un parque eólico empleando para diferentes horizontes de predicción $h$ . . . . .	101
4.16	Valor del CRPS para las diferentes distribuciones predictivas empleadas en los diferentes parques eólicos disponibles. . . . .	102

4.17	Comparación entre los percentiles obtenidos al evaluar la senda de densidades predictivas estimadas en 5 parques eólico empleando diferentes distribuciones . . . . .	103
5.1	Estimación adaptativa de un modelo TV-AAR(1) sobre la potencia ordenada de acuerdo al incremento del primer retardo . . . . .	108
5.2	Estimación adaptativa de un modelo TV-AAR(1) sobre la potencia ordenada de acuerdo a su primer retardo. . . . .	109
5.3	Estimación adaptativa de un modelo TV-AAR(1) sobre la potencia creciente ordenada de acuerdo a su primer retardo . . . . .	110
5.4	Estructura SETAR con 6 regímenes detectada en potencia eólica. . . . .	110
5.5	Graficos de caja para los umbrales detectados en los 76 ficheros de datos disponibles . . . . .	111



# Lista de tablas

3.1	Medidas para comparar los límites empíricos con los asintóticos empleando 1000 réplicas del modelo 3.27 para diferentes tamaños muestrales . .	60
3.2	Tasa de detección para diferentes contrastes en datos generados empleando el modelo (3.27) para diferentes valores de $\delta$ . . . . .	61
3.3	Tasa de detección obtenida empleando ARLS y el contraste de Tsay (1989) y el mejor de los resultados obtenido por los contrastes de Hansen (1996)	63
3.4	Modelos propuestos por Tsay, Tong y Aut-ARLS con sus respectivos AIC y BIC . . . . .	68
3.5	Modelos propuestos por Tsay, Tong y Aut-ARLS con sus respectivos AIC, error medio absoluto y raíz del error cuadrático medio. . . . .	70
3.6	Error cuadrático medio de los errores de predicción del PNB EE.UU.. . . .	72
5.1	Media de la raíz del error cuadrático medio para los 31 parques eólicos horarios. . . . .	118
5.2	Valor relativo de la raíz del error cuadrático medio obtenido para los 31 parques eólicos horarios. . . . .	118
5.3	Media del error absoluto medio para los 31 parques eólicos horarios. . . .	119
5.4	Valor relativo del error medio absoluto obtenido para los 31 parques eólicos horarios. . . . .	119
5.5	Media de la raíz del error cuadrático medio para los 19 parques eólicos 15 minutales. . . . .	120
5.6	Valor relativo de la raíz del error cuadrático medio obtenido para los 19 parques eólicos 15 minutales. . . . .	120
5.7	Media del error medio absoluto para los 19 parques eólicos 15 minutales.	121
5.8	Valor relativo del error medio absoluto obtenido para los 19 parques eólicos 15 minutales. . . . .	121
5.9	Media de la raíz del error cuadrático medio para los 21 aerogeneradores de un parque eólico con una frecuencia de 10 minutos. . . . .	122

5.10	Valor relativo de la raíz del error cuadrático medio de 21 aerogeneradores de un parque eólico con una frecuencia de 10 minutos. . . . .	123
5.11	Media del error absoluto medio para 21 aerogeneradores de un parque eólico con una frecuencia de 10 minutos. . . . .	123
5.12	Valor relativo del error medio absoluto para predicciones de 21 aerogeneradores de un parque eólico con una frecuencia de 10 minutos.. . . .	124



# Lista de acrónimos

**AIC** Criterio de selección de modelos de Akaike

**AAR** Modelo lineal autorregresivo ordenado

**AR** Modelo lineal autorregresivo

**ARCH** Modelo no lineal autorregresivo con heterocedasticidad condicional

**ARLS** Herramienta de identificación de modelos por umbrales propuesta, de nombre método de mínimos cuadrados recursivos y ordenados

**ARMA** Modelo lineal autorregresivo con media móvil

**Aut-ARLS** Procedimiento automático de la herramienta de identificación ARLS

**BIC** Criterio de selección bayesiano de modelos

**BL** Modelo no lineal bilineal

**BS** Brier Score, medida de evaluación de estimación de densidades

**CRPS** Criterio de selección de densidades (*Continuous Ranked Probability Score*)

**EAM** *Error medio absoluto*

**EAR** *Modelo no lineal autorregresivo exponencial*

**ECM** *Error cuadrático medio*

**EWMA** *Método de media móvil exponencialmente ponderada*

**EXPAR** *Modelo no lineal autorregresivo exponencial de amplitud dependiente*

**FAR** *Modelo no paramétrico autorregresivo con coeficientes funcionales*

**GARCH** *Modelo no lineal autorregresivo generalizado con heterocedasticidad condicional*

**MA** *Modelo lineal de media móvil*

**MEP** *Distribución de Máxima Entropía*

**MSAR** *Modelo no lineal autorregresivo con cambio de régimen dependiente de una cadena de Markov*

**NC** *Distribución Normal Censurada*

**NT** *Distribución Normal Truncada*

**OLS** *Método de estimación de mínimos cuadrados ordinarios*

**PIT** *Criterio de evaluación de estimación de densidades (Probability Integral Transform)*

**PNB** *Producto nacional bruto*

**RECM** *Raíz del error cuadrático medio*

**RLS** *Método de estimación de mínimos cuadrados recursivos*

**SETAR** *Modelo no lineal autorregresivo con umbrales auto-excitados*

**SETAR-TVCD** *Modelo no lineal autorregresivo con umbrales auto-excitados con distribución condicionada variable en el tiempo*

**STAR** *Modelo no lineal autorregresivo con umbrales suaves*

**TAR** *Modelo no lineal autorregresivo por umbrales*

**TV-AAR** *Modelo autorregresivo ordenado con parámetros variables en el tiempo*

**TV-AR** *Modelo autorregresivo con parámetros variables en el tiempo*

**WLS** *Método de estimación de mínimos cuadrados ponderados*

# Capítulo 1

## Introducción

Moran (1953) al realizar un ajuste lineal sobre el número de lince capturados anualmente en el distrito canadiense de *Mackenzie River* descubrió un, como él lo llamó, fenómeno curioso. Dicho fenómeno consistía en que los residuos para las observaciones mayores que la media eran significativamente menores que los residuos para aquellas observaciones que eran menores que la media. Seguramente nos encontramos ante una de las primeras identificaciones de la existencia de heterocedasticidad condicional en una serie temporal de datos reales.

Poco después, Whittle (1954) realizó un estudio sobre las mediciones del nivel del agua cada 15 segundos en un canal de la Isla de Bay en Nueva Zelanda. Analizando la estimación de la función de densidad espectral, Whittle descubrió la existencia de una relación aritmética que permitía relacionar los periodos de los picos existentes en dicha estimación. Whittle explica que la relación encontrada es debida a la existencia de no linealidad, realizando un análisis basado en ecuaciones diferenciales lineales a trozos. Tong (2010, p. 5) afirma que el trabajo de Whittle es, seguramente, el primer estudio acerca de la importancia de la idea del umbral en el análisis de las series temporales.

Motivados por trabajos similares a los de Moran y Whittle, el interés de investigadores por el estudio de técnicas no lineales en series temporales fue incrementándose de manera progresiva hasta nuestros días. Realmente, el primer estudio teórico de un modelo temporal no lineal fue efectuado por Volterra (1930). En su estudio mostró como cualquier función continua no lineal puede ser ajustada mediante la expansión de Volterra, cuya forma viene definida por

$$\begin{aligned} Y_t = & \sum_{u=0}^{\infty} g_u U_{t-u} + \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} g_{uv} U_{t-u} V_{t-v} \\ & + \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} \sum_{w=0}^{\infty} g_{uvw} U_{t-u} V_{t-v} W_{t-w} + \dots \end{aligned} \quad (1.1)$$

El interés de la formulación (1.1) es teórico únicamente, pues su estimación no es factible salvo que se impongan muchas restricciones.

La paradoja del análisis de series temporales no lineales es que, habitualmente, los fenómenos no lineales son más complejos de identificar y de modelizar que los fenómenos lineales, sin embargo, las herramientas disponibles son menos comprensibles y efectivas. El objetivo de esta tesis será desarrollar herramientas que faciliten el análisis de series temporales no lineales.

## 1.1 Modelos no lineales para series temporales

A lo largo de los años se han ido proponiendo diferentes tipos de modelos que intentan explicar diferentes fenómenos no lineales. En esta sección se hace un repaso de los modelos paramétricos habitualmente más empleados, así como una introducción a distintas técnicas no paramétricas existentes. Además de los modelos aquí mostrados existen otros más generales como pueden ser, por ejemplo, los modelos estado-dependientes (Priestley, 1980) o los modelos doblemente estocásticos (Tjøstheim, 1986).

### 1.1.1 Modelos por umbrales

Los modelos por umbrales, mencionados por vez primera por Tong (1977), asumen la existencia de diferentes funciones lineales en diferentes regiones del espacio de los estados. La división del espacio de los estados viene dada por la variable umbral  $X_{t-d}$ , con  $d \geq 1$ . Basándose en esta idea Tong (1987) formuló el principio del umbral, el cual dice que "el análisis de un sistema estocástico complejo puede ser realizado mediante su descomposición en subsistemas más sencillos" (Tong 1990, p. 99).

El modelo por umbrales más conocido es el autorregresivo por umbrales (TAR, *Threshold Autoregressive*) que tiene la forma

$$Y_t = a_0^{(J_t)} + \sum_{i=1}^p a_i^{(J_t)} Y_{t-i} + b^{(J_t)} e_t, \quad (1.2)$$

donde los  $\{e_t\}$  son variables aleatorias independientes e idénticamente distribuidas con  $E[e_t] = 0$  y  $V[e_t] = \sigma < \infty$ .  $\{J_t\}$  es un indicador temporal que toma valores en  $\{1, 2, \dots, J\}$  trabajando como mecanismo de intercambio. Dicho mecanismo es, habitualmente, una variable observada llamada variable umbral. La figura 1.1 muestra un ejemplo muy evidente de un modelo TAR con dos regímenes de comportamiento.

Los modelos TAR han sido empleado en diversas áreas, como pueden ser en ecología (p.ej., Stenseth et al., 1999), en hidrología (p.ej., Tong et al., 1985), en epidemiología (p.ej., Stenseth et al., 2006; Samia et al., 2007), en economía (p.ej., Tiao and Tsay, 1994; Hansen,

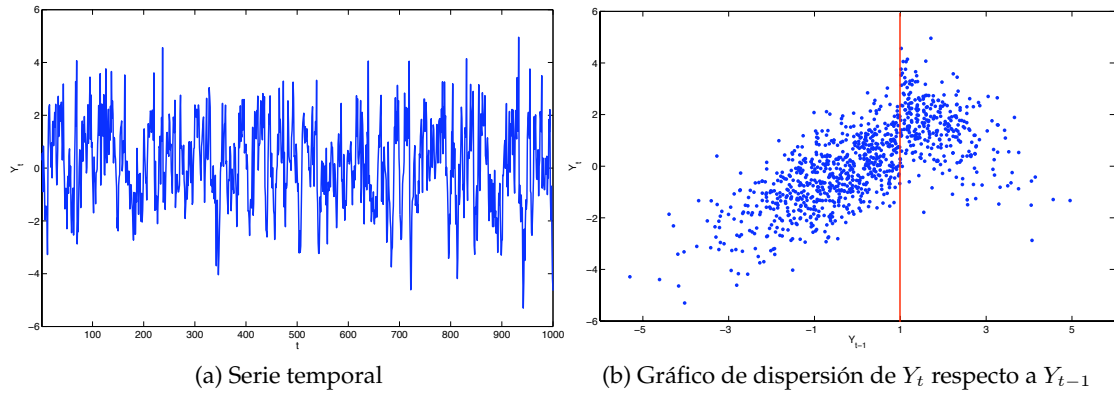


Figura 1.1: Realización de tamaño 1000 a partir del modelo (1.2) para valores  $J = 2$ ,  $a_0^{(1)} = 0$ ,  $a_0^{(2)} = 3$ ,  $a_1^{(1)} = .8$ ,  $a_1^{(2)} = -.8$ , dependiendo el indicador temporal  $J_t$  del valor de  $Y_{t-1}$ .

1999), en finanzas (p.ej., Li and Lam, 1995; Potter, 1995), en ciencias actuariales (p.ej., Chan et al., 2004), entre muchos otros. Un amplio estudio sobre el desarrollo de los modelos TAR ha sido efectuado por Tong (2010).

### 1.1.2 Modelos autorregresivos exponenciales de amplitud dependiente

Los modelos EXPAR (*Amplitude-dependent Exponential Autoregressive*) fueron introducidos de manera independiente por Jones (1976) y Ozaki y Oda (1978). Un modelo EXPAR(k) está definido por

$$Y_t = \sum_{j=1}^k [\alpha_j + \beta_j \exp(-\delta Y_{t-1}^2)] Y_{t-j} + e_t, \quad \delta > 0, \quad (1.3)$$

donde  $\{e_t\}$  es una secuencia de variables aleatorias independientes e idénticamente distribuidas. Haggan y Ozaki (1981) mostraron que el modelo es útil para la modelización de la vibración del sonido. En la figura 1.2 vemos un ejemplo de modelo EXPAR.

### 1.1.3 Modelos bilineales

Un proceso  $Y_t$  sigue un modelo Bilineal (BL(p,q,P,Q)) si está definido por

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + e_t + \sum_{j=1}^q \theta_j e_{t-j} + \sum_{i=1}^P \sum_{j=1}^Q \gamma_{ij} Y_{t-i} e_{t-j}, \quad (1.4)$$

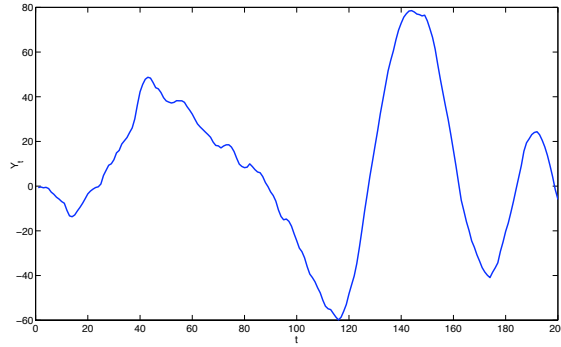


Figura 1.2: Realización de tamaño 200 del modelo (1.3), con  $k = 2$ ,  $\alpha = (2, -1)$ ,  $\beta = (.5, -.5)$  y  $\delta = 1$ .

y fueron introducidos por Granger y Andersen (1978). Son especialmente útiles para series temporales en las que ocasionalmente se produzcan grandes alejamientos de la media, como puede verse en la figura 1.3. Ejemplos de posibles aplicaciones reales podrían ser las mediciones realizadas por un sismógrafo, ya que éstas, en general, serán estables salvo cuando se produzca un terremoto. Subba Rao y Gabr (1984) hicieron un amplio estudio sobre este tipo de modelos.

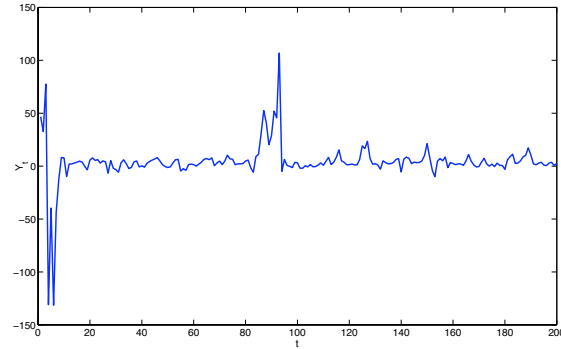


Figura 1.3: Realización de tamaño 200 del modelo (1.4), con  $p = q = P = Q = 1$ ,  $\phi = .8$ ,  $\theta = -.5$  y  $\gamma = .5$ .

#### 1.1.4 Modelos de heterocedasticidad condicional

Los modelos de heterocedasticidad condicional (ARCH, *Autoregressive Conditional Heteroscedastic*) fueron introducidos por Engel (1982) para modelizar la tasa de inflación de Reino Unido. Son modelos muy empleados en series temporales económicas y financieras. Un proceso  $Y_t$  sigue un modelo ARCH(r) si está definido como

$$Y_t = \sigma_t \varepsilon_t, \quad (1.5)$$

donde  $\{\varepsilon_t\}$  es un proceso de variables aleatorias independientes e idénticamente distribuidas, siendo  $\{\varepsilon_t\}$  independiente de  $\{Y_{t-k}, k \geq 1\}$ . La varianza seguirá un proceso autorregresivo de orden  $r$  definido por

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \dots + \alpha_r Y_{t-r}^2, \quad (1.6)$$

donde  $\sum_{i=1}^r \alpha_i < 1$ . La práctica ha demostrado que muchas series requieren de valores de  $r$  muy altos. Para evitar éste efecto, Bollerslev (1986) y Taylor (1987) propusieron un modelo más parsimonioso, los modelos autorregresivos generalizados de heterocedasticidad condicional (GARCH, *Generalized Autoregressive Conditional Heteroscedastic*). Un proceso  $Y_t$  sigue un modelo GARCH(p,q) si el proceso  $\sigma_t$  en (1.5) está definido como

$$\sigma_t^2 = c_0 + \sum_{i=1}^p c_i z_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2. \quad (1.7)$$

Los modelos GARCH tienen una estructura similar a los modelos ARMA, compartiendo con ellos muchas de sus propiedades. En la figura 1.4 se pueden ver dos realizaciones generadas a partir de modelos ARCH y GARCH. En el trabajo de De Gooijer y Hyndman (2006) se puede encontrar un resumen reciente de los diferentes estudios basados en modelos de heterocedasticidad condicional que se han ido proponiendo.

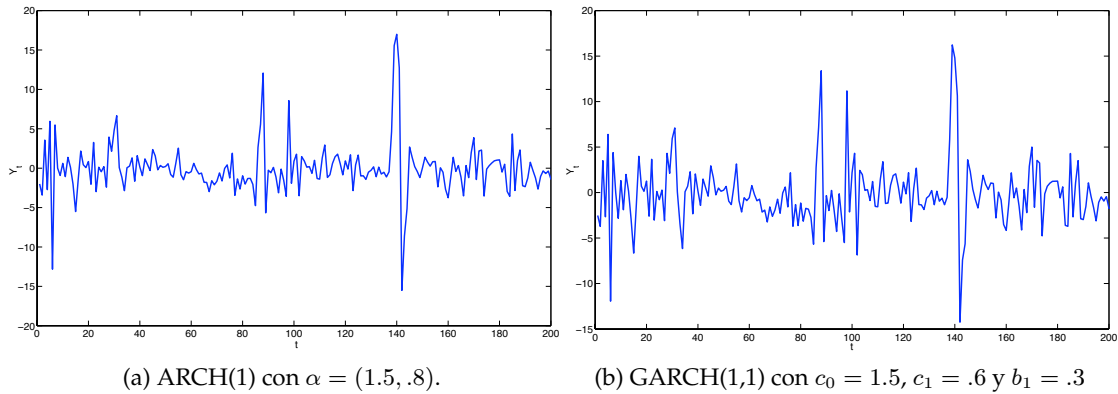


Figura 1.4: Ejemplos de modelos con heterocedasticidad condicional

### 1.1.5 Modelos de segunda generación

Los modelos mostrados hasta ahora intentan explicar un tipo de fenómeno no lineal, por ello Tong (1990, p.116) los llama modelos no lineales de primera generación. Para aprovechar mejor sus ventajas, Tong propone crear modelos más complejos que aúnen

a varios de estos modelos, creando así modelos de segunda generación, tercera generación, y así.

Un posible modelo de segunda generación que empieza a ser empleado es un modelo por umbrales con heterocedasticidad condicional. Por ejemplo, un modelo SETAR-ARCH vendría definido por

$$Y_t = a_0^{(J_t)} + \sum_{i=1}^p a_i^{(J_t)} Y_{t-i} + e_t V_t,$$

donde los  $\{e_t\}$  son variables aleatorias independientes e idénticamente distribuidas, y  $\{J_t\}$  es un indicador temporal que tomará valores en  $\{1, 2, \dots, J\}$  trabajando como mecanismo de intercambio. Además  $V_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \dots + \alpha_r Y_{t-r}^2$ . Este modelo combinará las ventajas de modelizar la media condicionada empleando un modelo lineal a trozos y la varianza condicionada a un proceso de heterocedasticidad condicional.

### 1.1.6 Modelos no paramétricos

En ocasiones los fenómenos no lineales subyacentes en los datos no pueden ser identificados de manera particular, por ello será adecuado el uso de modelos no paramétricos que permitan un ajuste no lineal más general. La forma de un modelo no paramétrico será

$$Y_t = f(Y_{t-1}, \dots, Y_{t-p}) + \sigma(Y_{t-1}, \dots, Y_{t-p}) e_t, \quad (1.8)$$

donde  $f(\cdot)$  y  $\sigma(\cdot)$  son funciones desconocidas y  $\{e_t\}$  es un proceso de variables aleatorias independientes e idénticamente distribuidas. El modelo (1.8) es conocido como proceso no paramétrico autorregresivo de heterocedasticidad condicional (NARCH, *Nonparametric Autoregressive Conditional Heteroscedastic*) o proceso no paramétrico autorregresivo (NAR) en el caso de que  $\sigma(\cdot)$  sea constante. El problema de este tipo de modelos es que para un valor de  $p > 2$ , la estimación de las funciones  $f(\cdot)$  y  $\sigma(\cdot)$  será prácticamente imposible, salvo que el tamaño muestral sea muy grande (ver, p.ej., Fan y Gijbels (1996)), debido a la maldición de la dimensionalidad (Bellman, 1961).

Debido al problema de la dimensionalidad para trabajar con modelos no paramétricos hay dos opciones. Por una parte existen diferentes técnicas que permiten estimar las funciones  $f(\cdot)$  y  $\sigma(\cdot)$  empleando métodos kernel, splines,... (ver, p.ej., Fan y Yao, 2005 Caps. 5 y 8; Peña et al., 2001 Cap. 12). Otra opción diferente es emplear modelos para ajustar dichas funciones. A continuación vemos algunos de estos modelos.



### Modelo de coeficientes funcionales

El modelo de coeficientes funcionales (FAR, *Functional-Coefficient Autoregressive*), introducido por Chen y Tsay (1993), tiene la forma

$$Y_t = f_1(Y_{t-d})Y_1 + \dots + f_p(Y_{t-d})Y_{t-p} + \sigma(Y_{t-d})e_t, \quad (1.9)$$

donde  $\{e_t\}$  es una secuencia de variables aleatorias independientes e idénticamente distribuidas, y además  $\{e_t\}$  será independiente de  $\{Y_{t-k}, k \geq 1\}$ . Los coeficientes funcionales  $f_1(\cdot), \dots, f_p(\cdot)$  son desconocidos. El modelo (1.9) depende de la variable dependiente  $Y_{t-d}$ , que es un retardo del proceso, limitando el alcance de las aplicaciones posibles. Por ello, una generalización de esta clase de modelos es permitir una combinación lineal de valores pasados como variable dependiente del modelo. Esta generalización es conocida como modelo de coeficientes funcionales adaptativo.

### Modelos aditivos

Los modelos aditivos (Ezekiel, 1924) son muy útiles para aproximar la función autorregresiva de alta dimensión del modelo (1.8) mediante la descomposición de la función  $f(Y_{t-1}, \dots, Y_{t-p})$  en la suma de funciones univariantes, es decir, un modelo aditivo tiene la forma

$$Y_t = f_1(Y_{t-1}) + \dots + f_p(Y_{t-p}) + e_t, \quad (1.10)$$

donde  $\{e_t\}$  es una secuencia de variables aleatorias independientes e idénticamente distribuidas. Las funciones  $f_1, \dots, f_p$  son univariantes y pueden ser estimadas mediante alguna técnica de regresión no paramétrica. Lógicamente, esto disminuye la dimensión del problema, y nos da la ventaja de necesitar menos observaciones para obtener resultados precisos. El problema es que la aditividad es una condición muy fuerte, que habitualmente no se cumple, y que además es difícil de comprobar. En Chen et al. (1995) se propone un contraste que permite evaluar si se cumple la condición de aditividad.

## 1.2 Problemas tratados en la tesis

En esta sección se describen los problemas a los que nos hemos enfrentado durante la realización de esta tesis, así como las propuestas que más adelante serán desarrolladas.

### 1.2.1 Identificación de modelos no lineales

El uso de un modelo lineal para explicar un conjunto de datos que presenten algún tipo de fenómeno no lineal, provocará que el modelo no pueda recoger toda la informa-

ción contenida en los datos. En la figura 1.5 vemos un ejemplo de un proceso temporal  $Y_t$  cuya relación con su primer retardo  $Y_{t-1}$  es no lineal. Como vemos, si realizamos un ajuste lineal será complicado que se pueda predecir su comportamiento futuro adecuadamente. Por ello, es importante identificar cuándo tenemos algún fenómeno no lineal subyacente en los datos para así poder aplicar métodos más adecuados. El problema radica en cómo identificar estos fenómenos, cómo identificar la no linealidad.

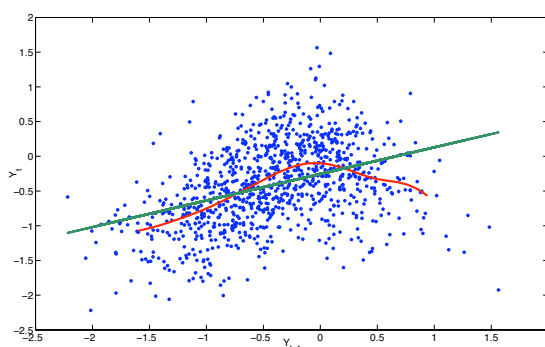


Figura 1.5: Gráfico de dispersión de  $Y_t$  respecto a  $Y_{t-1}$ .

Identificar un fenómeno no lineal particular suele ser una tarea compleja. Para ello se pueden emplear diferentes métodos como pueden ser análisis de los datos mediante métodos gráficos, realización de contrastes de linealidad

### Métodos gráficos

El uso de métodos gráficos nos puede mostrar que algo anormal pasa en nuestros datos. Sin duda, son una herramienta muy útil cuando se comienza a analizar cualquier tipo de datos. Métodos como histogramas, correlogramas o funciones de densidad espectral, que son habitualmente empleados en los análisis lineales puede darnos una idea de que algo no funciona bien, en caso de observar fenómenos no esperados. A dichas herramientas gráficas añadimos aquí algunas propuestas realizadas en el ámbito de las series temporales no lineales.

**Gráfico inverso** Una serie temporal estacionaria  $\{Y_t\}$  es reversible en el tiempo si para cada entero positivo  $n$ , y para cada vector de índices  $(t_1, \dots, t_n)$  los vectores  $(Y_{t_1}, \dots, Y_{t_n})$  y  $(Y_{-t_1}, \dots, Y_{-t_n})$  tienen la misma distribución conjunta. Tong (1990, p.197) explica cómo la falta de reversibilidad en el tiempo indica falta de linealidad en los datos. El problema es que existen modelos no lineales que pueden cumplir la propiedad de reversibilidad por lo que dicha propiedad no puede ser empleada para descartar que un proceso es no lineal.

Para comprobar que se cumple la propiedad de reversibilidad en el tiempo, Tong (1990, p.216) propone emplear los gráficos inversos. Dichos gráficos fueron propuestos por Cleveland y McGill (1987) para poder detectar diferentes características interesantes en los datos. En la figura 1.6 podemos ver ejemplos de procesos reversibles y no reversibles en el tiempo.

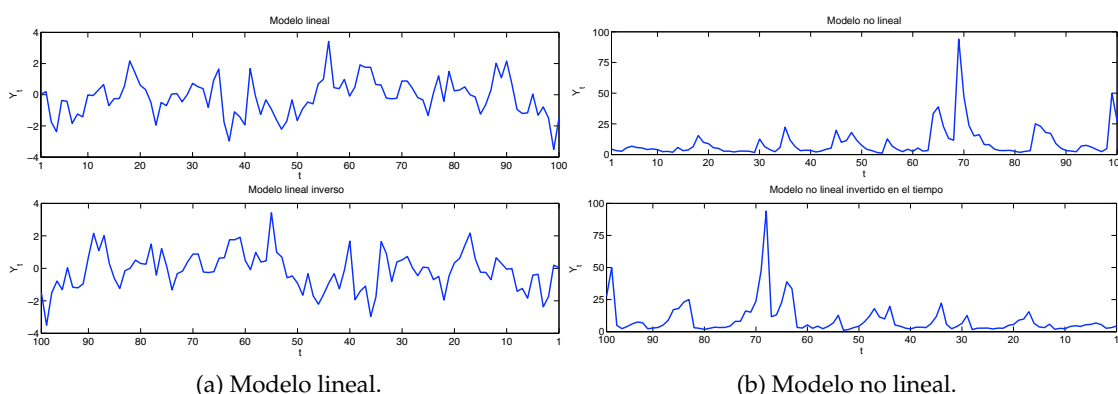


Figura 1.6: Comparación del gráfico de una serie procedente de modelos lineales y no lineales

**Diagrama de dispersión directo** Se trata de gráficos de dispersión del proceso  $\{Y_t\}$  respecto a su retardo  $\{Y_{t-k}\}$ , en el cual los puntos consecutivos se van uniendo mediante líneas. En la figura 1.7 vemos un ejemplo del diagrama para datos generados mediante un modelo AR lineal y un modelo SETAR no lineal. El diagrama perteneciente al proceso lineal tiene un comportamiento completamente aleatorio, mientras que el proceso no lineal dibuja una especie de elipse, sin datos en el centro. Tong (1990, p.216) explica cómo este tipo de resultados muestran la existencia de ciclos. Este tipo de gráficos ha sido muy empleado en el análisis de técnicas de caos.

**Suavizado no paramétrico** El estudio del suavizado no paramétrico realizado sobre la relación existente proceso  $\{Y_t\}$  respecto a su retardo  $\{Y_{t-k}\}$  puede ser una herramienta útil para identificar una posible relación no lineal. En la figura 1.8 vemos la diferencia entre realizar un ajuste lineal y un suavizado no paramétrico, y cómo este último ayuda a explicar mejor los datos, pudiendo indicar por lo tanto no linealidad.

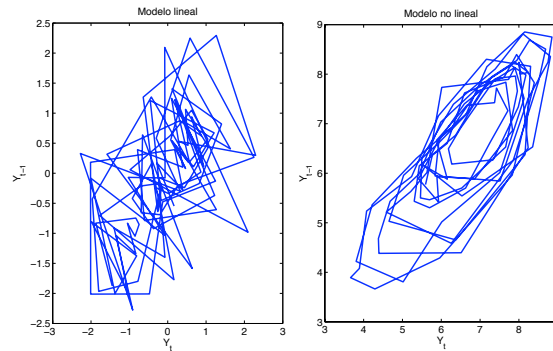


Figura 1.7: Diagramas de dispersión directos para datos provenientes de modelos lineales y no lineales

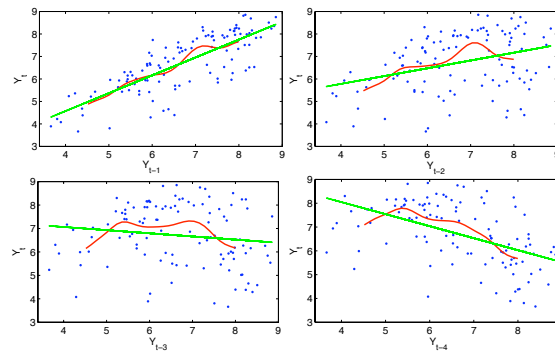


Figura 1.8: Ajuste lineal y suavizado no paramétrico sobre un proceso generado a partir de un modelo no lineal

### Contrastes de linealidad

Una herramienta más formal para comprobar la existencia de no linealidad en el proceso es el uso de contrastes de hipótesis. Existen dos clases de contrastes de linealidad, por una parte están los contrastes generales que no asumen ningún tipo de estructura no lineal en los datos, sino que solo contrastan la falta de linealidad en los mismos. Por otra parte, están aquellos contrastes específicos que nos permiten comprobar si un tipo de fenómeno no lineal en particular es correcto para el proceso estudiado.

**Contrastes generales** La hipótesis a contrastar es, en general, se contrasta si la mejor representación lineal del proceso es correcta o no, para ello existen multitud de contrastes. Uno de los primeros que se propusieron es el contraste basado en la aproximación bi-espectral, propuesto originalmente por Subba Rao y Gabr (1980) y mejorado poco después por Hinich (1982). Se basa en la idea de que bajo el supuesto de linealidad las funciones de densidad espectrales acumuladas son constantes. Su principal

problema es la estimación de las densidades espectrales.

Ramsay (1969) desarrolló un contraste de linealidad para modelos de regresión basado en el contraste de no aditividad de Tukey (1949). Usando esta idea, Keenan (1985) diseñó un contraste de no linealidad en series temporales. Dicho contraste fue mejorado por Tsay (1986).

Otro tipo de contrastes generales se basan en la idea propuesta por Granger y Andersen (1978). En su trabajo mostraron que sea  $Y_t$  es un proceso gaussiano estacionario

$$\rho_j(Y_t^2) = \{\rho_j(Y_t)\}^2, \text{ para todo } j, \quad (1.11)$$

donde  $\rho_j(Y_t)$  es la función de autocorrelación del proceso  $\{Y_t\}$ . Según Granger y Andersen el no cumplimiento de esta propiedad podría indicar falta de linealidad en los datos. Maravall (1983) mostró que los residuos resultantes de realizar un ajuste lineal en el proceso han de cumplir la relación (1.11), ya que en caso contrario el proceso será no lineal. Empleando esta idea, McLeod y Li (1983) propusieron un contraste de Ljung-Box que permite comprobar que la relación (1.11) se mantiene en los residuos. Curiosamente, aunque McLeod y Li (1983) realizaron un contraste general, su propuesta funciona especialmente bien para detectar la existencia de modelos de heterocedasticidad condicional. Siguiendo esta misma vía, Peña y Rodríguez (2005) proponen un contraste que mejora a los anteriores.

Otros contrastes de linealidad existentes son el contraste BDS propuesto por Brock et al. (1996) que se basa en la relación no lineal de los residuos del ajuste lineal y de su pasado, y el contraste del tipo Kolmogorov-Smirnov propuesto por Hong-zhi y Bing (1991).

**Contrastes específicos** Los contrastes específicos tienen como hipótesis alternativa la existencia de un modelo no lineal. Repasamos brevemente algunas propuestas de los modelos más empleados habitualmente.

Para contrastar la existencia de modelos por umbrales existen diferentes tipos de propuestas. Los más empleados son contrastes de ratio de verosimilitud (Chan y Tong, 1990; Chan, 1990; Hansen, 1996, 1999; Ling y Tong, 2005) y contrastes basados en los residuos predictivos (Petrucelli y Davies, 1986; Tsay, 1989).

Para detectar la existencia de heterocedasticidad condicional son muy útiles los contrastes basados en los residuos al cuadrado, como dijimos anteriormente. El problema de estos es que simplemente dan un indicio de existencia de heterocedasticidad condicional en los residuos. Para comprobar de manera más específica la posible existencia de efectos ARCH o GARCH, existen contrastes como los de Lee y King (1993) o Hong (1997).

Fan y Yao (2005, Cap. 9) desarrollan un contraste de ratio de verosimilitud generalizado basado en el contraste para datos independientes de Fan et al. (2001). La idea principal de este contraste es que la distribución asintótica del estimador bajo la hipótesis nula es independiente de los parámetros desconocidos ausentes de dicha hipótesis nula. Esta propiedad es conocida como fenómeno de Wilks. Fan y Yao (2005) muestran como siguiendo esta idea se puede desarrollar un contraste para series temporales, que permite contrastar la existencia de diferentes modelos no paramétricos o paramétricos, como el TAR, el EXPAR, el BL, el FAR, etc.

Por último, existen diferentes contrastes de bondad de ajuste que permiten comparar diferentes modelos tanto lineales como no lineales entre sí, como son las propuestas de Escanciano (2006), Ling y Tong (2010) o Du y Escanciano (2011).

### Resultados propuestos en la tesis

En el análisis de las series temporales lineales existe una medida de dependencia general que permite la modelización del proceso, se trata de la autocorrelación simple y parcial. Además, esta medida permite una sencilla representación gráfica mediante el correlograma. Sin embargo, no existe ninguna medida ni ninguna herramienta, que permita de manera tan simple e ilustrativa analizar las relaciones no lineales. Por ello, el primer objetivo de esta tesis doctoral es el desarrollar una herramienta de identificación y modelización de un fenómeno no lineal. En particular, durante esta tesis nos hemos centrado en la identificación de modelos por umbrales.

Para ello vamos a proponer un procedimiento basado en una herramienta gráfica que permita identificar la existencia de un modelo por umbrales. Además, dicho procedimiento nos permitirá modelizar de manera sencilla el proceso.

#### 1.2.2 Predicción con modelos no lineales

El cálculo de predicciones es una de las tareas esenciales cuando se trabaja con series temporales. Existen dos tipos de caminos que se pueden emplear para calcular predicciones, por una parte el uso de modelos que expliquen el comportamiento de los datos, permitirá además obtener predicciones a partir de ellos. Mientras que otra opción es el uso de técnicas de predicción no paramétricas que sin explicar el comportamiento de la serie permitirán calcular predicciones. En esta tesis nos hemos centrado en el cálculo de predicciones empleando modelos no lineales.

Sea  $Y_t$  un proceso conocido, el objetivo será calcular predicciones a un horizonte  $h$  dada la información contenida en dicho proceso mediante

$$\hat{Y}_{t+h|t} = f(Y_t, \dots, Y_1), \quad (1.12)$$

donde la función  $f(\cdot)$ , si asumimos una función de pérdida cuadrática será la esperanza condicionada (ver, p.ej. Chatfield 2001), es decir,

$$\hat{Y}_{t+h|t} = E_t [Y_{t+h} | Y_t, \dots, Y_1]. \quad (1.13)$$

Si el modelo ajustado es no lineal la forma analítica de la esperanza condicionada será muy complicada de calcular. Para estos casos, Tong (1990) propone estimar la densidad de las predicciones condicionada a la información conocida en el instante de tiempo  $t$  y a partir de ella calcular los momentos condicionados. A dicha densidad de las predicciones se le conoce habitualmente como densidad predictiva. Cuando trabajamos con modelos no lineales, lo habitual será que el proceso no sea gaussiano, por lo que el cálculo de la densidad predictiva será complejo. Un estudio más profundo acerca de la predicción mediante modelos no lineales y mediante técnicas no paramétricas puede encontrarse en Tong (1990, Cap.6) y Fan y Yao (2005, Cap. 10).

### Predicciones puntuales con modelos TAR

Un modelo por umbrales está compuesto de modelos lineales, entre los que el proceso irá cambiando dependiendo de la variable umbral. Debido a la linealidad de los diferentes regímenes el uso de modelos TAR será ventajoso respecto a otros modelos no lineales a la hora de calcular predicciones. Su principal problema surge cuando el horizonte de predicción  $h$  es mayor que el retardo de la variable umbral  $Y_{t-d}$ . En ese caso, la variable umbral no será observada, con lo que será complicado encontrar la forma analítica de las predicciones. Tong (1990) muestra como hacerlo mediante ecuaciones de Chapman-Kolmogorov las predicciones, lo que para estructuras TAR muy complejas puede resultar extremadamente dificultoso. Clements y Smith (1997) muestran diferentes métodos de computación que permiten el cálculo de predicciones en modelos TAR mediante simulaciones. Para realizar estas simulaciones será necesario estimar la densidad condicionada del proceso. En la sección 2.6 mostraremos más detenidamente las diferentes formas de calcular predicciones empleando un tipo de modelo TAR.

### 1.2.3 Aplicaciones en energía eólica

El uso de la energía eólica como fuente de energía renovable ha experimentado un gran incremento en la última década. Por ejemplo, como podemos ver en la figura 1.9a, la potencia instalada en España ha pasado de ser el 3.4% del total instalado en 2000 (1875MW) a ser el 19.4% (19959MW) en 2010. Como vemos en la figura 1.9a en apenas 10 años la potencia instalada correspondiente a las energías renovables ha pasado de estar por debajo del 40% a rondar el 50% del total. En particular vemos como la energía

eólica ha pasado de ser una energía residual dentro de las renovables a representar casi la mitad de la potencia instalada.

Con la demanda cubierta por las diferentes fuentes energéticas sucede algo semejante. Como vemos en la figura 1.9b, la potencia eólica adquirida en el mercado energético ha pasado de ser prácticamente inexistente a cubrir en el año 2010 el 15.6% del total demandado. Este aumento ha provocado la necesidad de obtener mejores predicciones tanto para poder optimizar su integración en el mercado energético como por motivos de operatividad del sistema eléctrico.

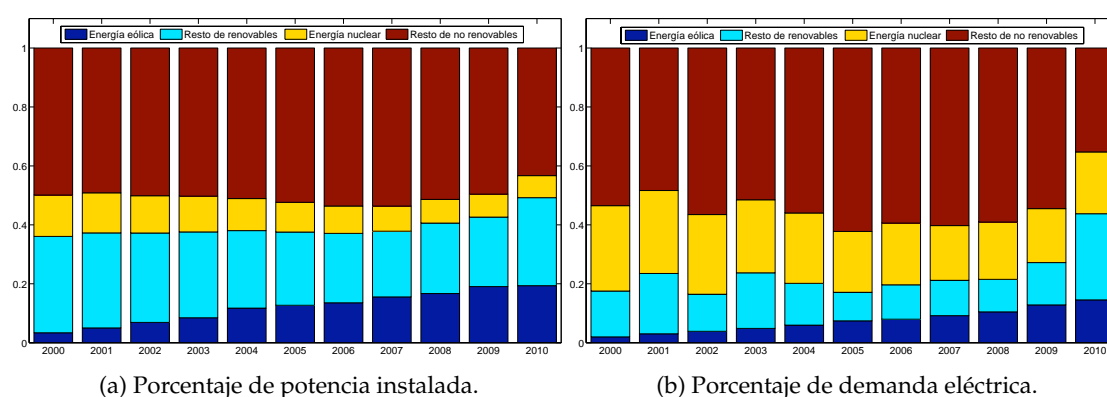


Figura 1.9: Evolución del sistema eléctrico español

### Propuesta de cálculo de predicciones puntuales

La desventaja de la energía eólica en el mercado eléctrico respecto a otras fuentes de energía es su incertidumbre. Para minimizar esta incertidumbre es necesario obtener predicciones que permitan su integración en el mercado eléctrico. Estas predicciones deben ser a muy corto plazo, debido a la existencia de un mercado eléctrico intradiario.

La realización de predicciones puntuales de potencia eólica empleando modelos estadísticos ha sido ampliamente estudiada en años recientes (Sánchez, 2006a; Costa et al., 2008; Giebel et al., 2011). El uso de un modelo que permite la existencia de diferentes regímenes de comportamiento entre los que irá cambiando dependiendo del valor de una variable observada puede ser muy útil para modelar la potencia eólica, puesto que es un proceso que cambia de manera abrupta como vemos en la figura 1.10. Por ello, proponemos emplear modelos por umbrales para realizar predicciones a corto plazo en potencia eólica, como también hiciera Pinson et al. (2008).



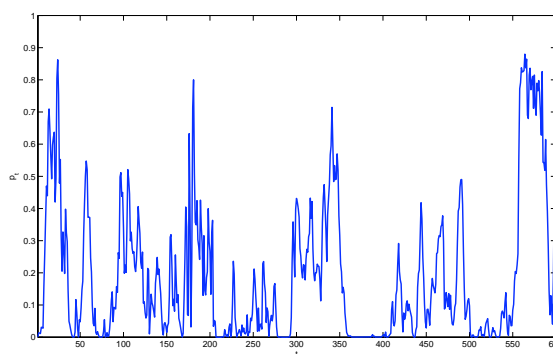


Figura 1.10: Ejemplo de serie de datos de producción eólica

### Propuesta de estimación de densidades condicionadas

Las predicciones puntuales no son suficientes, ya que en ocasiones se hace necesario obtener la estimación de la densidad predictiva, para así tener un mejor conocimiento acerca del comportamiento de las predicciones (Diebold y López, 1996). En los últimos años han sido múltiples las propuestas de métodos de estimación de densidades predictivas, como puede verse en Tay y Wallis (2000), Mitchell (2008) o Clements y Harvey (2010).

En particular, la utilización de funciones de densidad predictiva para la energía eólica permitirá, entre otras cosas, mejorar la gestión de las reservas de energía eléctrica, mejorar la programación de las centrales de energía eléctrica convencionales y minimizar el riesgo que cometen los agentes que ofrecen energía eólica en el mercado eléctrico.

El problema de estimación de densidades predictivas no ha sido muy tratado cuando trabajamos con la potencia eólica. En este caso la estimación es más compleja aún que en los casos tratados tradicionalmente, ya que la variable es no gaussiana, y además está acotada entre 0 y la potencia nominal. Como en general se trabaja estandarizando la potencia producida por la potencia nominal, los datos estarán acotados entre 0 y 1.

Entre las propuestas realizadas en la literatura, existen diferentes puntos de vista desde los que abordar el problema de la estimación de la densidad predictiva. Una posible opción es trabajar directamente con la distribución de la velocidad del viento para luego emplear la relación no lineal existente entre velocidad y potencia generada para poder estimar la distribución de la potencia (p.ej., Lange, 2005; Carta et al. 2009). Otra propuesta similar, aunque más compleja, se basa en la realización de múltiples predicciones de variable meteorológicas a partir de las cuales se estima la distribución de la producción (p.ej., Nielsen et al. 2004; Pinson y Madsen 2009; Taylor et al. 2009).

Por último otras propuestas diferentes se basan en la estimación de la densidad pre-

dictiva a partir de los datos históricos de la potencia y de sus predicciones (p.ej., Bludszuweit et al., 2008; Al-Awami y El-Sharkawi, 2009; Lau y McSharry, 2010; Pinson, 2010). Este punto de vista será el empleado en nuestra propuesta de estimación. Empleando las predicciones de potencia estimaremos de manera adaptativa los momentos condicionados de los errores de predicción. Con estos momentos condicionados, estimaremos la senda de predicciones predictiva ajustando diferentes distribuciones paramétricas.

### 1.3 Estructura de la tesis

La tesis consta de 6 capítulos. En este primer capítulo se han introducido los diferentes problemas que hemos tratado a lo largo de la tesis. Para ello realizaremos una descripción de diferentes técnicas existentes para analizar series temporales no lineales. Además se introducen los problemas básicos a los que nos hemos enfrentado en esta tesis como son, por una parte, la identificación de no linealidad en datos temporales, y por otra la estimación de densidades predictivas y el cálculo de predicciones puntuales empleando modelos no lineales. En particular, en estas últimas tareas nos concentraremos en su aplicación a datos provenientes de la potencia producida por parques eólicos, por lo que introduciremos diferentes técnicas existentes que se emplean para este tipo de datos.

A lo largo de esta tesis nos hemos centrado en un tipo de modelos no lineales, que son los modelos por umbrales, por ello en el capítulo 2 repasamos de manera exhaustiva los diferentes tipos de modelos por umbrales que se han propuesto. En particular, repasaremos los modelos autorregresivos por umbrales.

Los principales resultados de la tesis se encuentran en los capítulos 3, 4 y 5. En el capítulo 3 desarrollamos un procedimiento que permitirá identificar y modelizar los modelos autorregresivos por umbrales. Dicho procedimiento se basa en dos puntos principales, por una parte ordenamos los datos respecto a la variable que provoca la no linealidad, como ya hicieron en sus propuestas de contrastes Petrucci y Davies (1986) y Tsay (1989), y por otra parte, empleamos un método de estimación recursiva sobre los datos ordenados para así poder detectar posibles cambios estructurales provocados por la variable que empleamos para realizar la ordenación. A lo largo del capítulo mostraremos el método recursivo empleado así como sus principales propiedades y justificaremos porqué la ordenación no altera la dependencia temporal de la variable. Por último compararemos el procedimiento propuesto con diferentes contrastes existentes mostrando como los mejoraremos.

En el capítulo 4 se trata la estimación de la densidad predictiva, cuando la variable

empleada es la potencia generada por un parque eólico. Mostraremos de que manera las características principales de este tipo de datos que complican la estimación de la densidad predictiva, así como el método de estimación adaptativa mediante el que proponemos estimar los momentos centrales de las predicciones puntuales. Empleando estos momentos detallaremos como estimar la densidad predictiva empleando diferentes distribuciones. Por último, evaluaremos las estimaciones de las densidades predictivas.

La estimación de la densidad predictiva es empleada a lo largo del capítulo 5 para realizar estimaciones puntuales empleando un modelo no lineal. Es conocido, que el cálculo de modelos no lineales puede ser en ocasiones complejo. En particular, empleando modelos autorregresivos por umbrales a horizontes elevados es habitual emplear un procedimiento de Monte Carlo para calcular las predicciones. Para ello es necesario realizar una estimación de la densidad predictiva si queremos evitar el supuesto de normalidad en la variable, supuesto que en el caso de datos de potencia eólica no se cumple.

Por último en el capítulo 6, resumiremos las principales conclusiones así como las principales contribuciones contenidas en la tesis.



## Capítulo 2

# Modelos autorregresivos por umbrales

La forma general de un modelo por umbrales (TAR, *Threshold Autoregressive*) es

$$Y_t = a_0^{(J_t)} + \sum_{i=1}^p a_i^{(J_t)} Y_{t-i} + b^{(J_t)} e_t, \quad (2.1)$$

donde los  $\{e_t\}$  son variables aleatorias independientes e idénticamente distribuidas con  $E[e_t] = 0$  y  $V[e_t] = \sigma < \infty$ , y  $\{J_t\}$  es un indicador temporal que tomará valores en  $\{1, 2, \dots, k\}$ . La idea básica de un modelo por umbrales es la descomposición del sistema no lineal en  $k$  sistemas lineales más sencillos, mediante el indicador temporal  $\{J_t\}$ . Esta idea es el denominado principio del umbral (Tong, 1987). El modelo TAR puede ser fácilmente extendido añadiendo parte media móvil (TARMA, *Threshold Autoregressive Moving Average*), reemplazando en (2.1), la expresión  $b^{(J_t)} e_t$  por  $\sum_{j=0}^q b_j^{(J_t)} e_{t-j}$ .

Una de las claves del éxito de los modelos TAR está en su simplicidad para ser interpretado, al contrario que otros tipos de modelos no lineales. En este sentido, es similar a los modelos lineales. Sin embargo, el desarrollo teórico de los modelos TAR aún no ha sido completado. El capítulo comienza con una revisión de diferentes tipos de modelos por umbrales que se han ido proponiendo. A continuación, nos centraremos en el objetivo principal del capítulo, los modelos autorregresivos por umbrales autoalimentados.

## 2.1 Tipos de modelos por umbrales

Un modelo autorregresivo por umbrales auto-alimentado (SETAR( $k; [p_1, \dots, p_k]; d$ ), *Self-Exciting Threshold Autoregressive*) con  $k$  regímenes, siendo  $k \geq 2$ , se define como

$$Y_t = \sum_{i=1}^k \left( \phi_0^{(i)} + \phi_1^{(i)} Y_{t-1} + \dots + \phi_{p_i}^{(i)} Y_{t-p_i} + e_t^{(i)} \right) I(Y_{t-d} \in \Omega_i), \quad (2.2)$$

donde  $\Omega_i$  es una partición del dominio  $(-\infty, \infty)$  de manera que  $\cup_{i=1}^k \Omega_i = (-\infty, \infty)$  y  $\Omega_i \cap \Omega_j = \emptyset, \forall i \neq j$ . Las particiones del dominio están dictadas por el valor de la variable umbral  $Y_{t-d}$ , donde  $d$  ( $d \in \mathbb{Z}^+$ ) es conocido como parámetro *delay*, y se forman mediante  $\Omega_i = (r_{i-1}, r_i]$ . Los valores  $r_i$  son los umbrales y cumplen que  $-\infty = r_0 < r_1 < \dots < r_k = \infty$ . En cada régimen,  $\{e_t^{(i)}\}$  es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media cero y desviación finita  $\sigma_{(i)}$ . La función  $I(\cdot)$  es una función indicadora, que toma valor 1 cuando se cumple la condición y 0 en otro caso. El modelo SETAR es quizás el más conocido entre todos los modelos por umbrales que se han ido proponiendo. Su extensión directa será añadiendo el componente media móvil, con lo que un modelo SETARMA( $k; [p_1, \dots, p_k], [q_1, \dots, q_k]; d$ ) se define como

$$Y_t = \sum_{i=1}^k \left( \phi_0^{(i)} + \phi_1^{(i)} Y_{t-1} + \dots + \phi_{p_i}^{(i)} Y_{t-p_i} + e_t^{(i)} - \theta_1^{(i)} e_{t-1}^{(i)} - \dots - \theta_{q_i}^{(i)} e_{t-q_i}^{(i)} \right) I(Y_{t-d} \in \Omega_i). \quad (2.3)$$

Otro tipo de modelos posible es aquél en el que el proceso indicador  $\{J_t\}$  es independiente del proceso  $Y_t$ . En ese caso, la variable umbral será una variable exógena al proceso  $Y_t$ . Tong y Lim (1980) nombraron a este tipo de modelos con variables exógenas como modelos autorregresivos por umbrales con sistema de bucle abierto, pensando principalmente, en términos de control de sistemas. En su reciente estudio sobre la evolución de los modelos por umbrales Tong (2010) propone modificar el nombre a modelos TARMAX (*Threshold Autoregressive Moving Average Exogenous*). De esta manera  $Y_t$  seguirá un modelo TARMAX( $k; [p_1, \dots, p_k], [q_1, \dots, q_k]; d$ ) si está definido por

$$Y_t = \sum_{i=1}^k \left( \phi_0^{(i)} + \phi_1^{(i)} Y_{t-1} + \dots + \phi_{p_i}^{(i)} Y_{t-p_i} + e_t^{(i)} - \theta_1^{(i)} e_{t-1}^{(i)} - \dots - \theta_{q_i}^{(i)} e_{t-q_i}^{(i)} \right) I(X_{t-d} \in \Omega_i), \quad (2.4)$$

En los modelos vistos hasta ahora el cambio de región en el espacio de los estados viene dado por una función indicadora, lo que provocará un cambio brusco entre regímenes. En algunas aplicaciones, será más apropiado que el cambio entre re-

giones se produzca de manera más "suave", pudiendo usar como proceso indicador  $\{J_t\}$  una función suavizada. Los primeros en proponer este tipo de modelos fueron Chan y Tong (1986) llamándolos modelos autorregresivos con umbrales suaves (STAR, *Smooth Threshold Autoregressive*). Los modelos STAR han sido ampliamente estudiados en el campo de la econometría (p.ej., Franses y van Dijk, 2000), siendo habitualmente denominados como modelos autorregresivos con transición suave (*Smooth Transition Autoregressive*).

Todos los modelos anteriores emplean como variable umbral una variable observada, pero hay otro tipo de modelos en los que dicha variable umbral no será observada. El primero de ellos es el modelo autorregresivo exponencial (EAR, *Exponential Autoregressive*). Sea  $\{J_t\}$  una secuencia de variables aleatorias independientes e idénticamente distribuidas tal que

$$J_t = \begin{cases} 1, & \text{con probabilidad } 1 - \alpha \\ 2, & \text{con probabilidad } \alpha \end{cases}, \quad (2.5)$$

entonces el modelo EAR(2) tendrá la forma

$$Y_t = \begin{cases} a_1^{(1)} Y_{t-1} + e_t, & \text{con probabilidad } 1 - \alpha \\ a_2^{(2)} Y_{t-1} + e_t, & \text{con probabilidad } \alpha \end{cases},$$

que como puede verse es el modelo TAR (2.1) cuando  $a_0^{(J_t)} = 0$ ,  $b^{(J_t)} = 1$ ,  $0 \leq a_1^{(1)} < 1$ ,  $a_1^{(2)} = 0$ ,  $a_2^{(1)} = 0$  y  $0 \leq a_2^{(2)} < 1$ . El modelo EAR fue introducido por Gaver y Lewis (1980) y Lawrance y Lewis (1980). La conexión entre el modelo EAR y el modelo TAR fue notada por vez primera por Tong (1983, p.63) y posteriormente fue desarrollada por Chan (1986, Cap.4; 1988) extendiendo el modelo EAR a ordenes más elevados.

Otro tipo de modelos en los que la variable umbral es no observada, es el modelo de Markov con intercambio oculto (MSAR, *Markov Switching Autoregressive*). En este caso, estamos ante el modelo TAR definido en (2.1) siendo el indicador temporal  $\{J_t\}$  una cadena de Markov finita. Este modelo fue propuesto por primera vez por Tong y Lim (1980, p.285), siendo probablemente el trabajo de Tyssedal y Tjøstheim (1988) el primer estudio en profundidad que se realizó. Una posible modificación del modelo MSAR es que el indicador temporal  $\{J_t\}$  sea parcialmente oculto, como propuso Tong (1983, p. 276-277). Esta modificación solo ha sido explorada parcialmente, siendo la propuesta de Wu y Chen (2007) de las más recientes.

## 2.2 Estacionariedad en modelos SETAR

Para el estudio de la estacionariedad en modelos SETAR se reescribe la expresión (2.2) en forma de un modelo general autorregresivo no lineal

$$Y_t = h(Y_{t-1}, \dots, Y_{t-p}) + e_t, \quad (2.6)$$

siendo  $h$  una función lineal a trozos, y donde el proceso  $\{e_t\}$  es una secuencia de variables aleatorias independientes e idénticamente distribuidas. Al ser el proceso  $Y_t$  una serie temporal, se asume de manera implícita que  $e_t$  es independiente de  $Y_{t-j}$ , para  $j \geq 1$ . Ahora, sean  $\mathbf{Y}_t = (Y_t, \dots, Y_{t-p+1})'$  y  $\mathbf{e}_t = (e_t, 0, \dots, 0)'$ , entonces, a partir del modelo (2.6), se deduce que  $\mathbf{Y}_t$  es una cadena de Markov definida como

$$\mathbf{Y}_t = \mathbf{f}(\mathbf{Y}_{t-1}) + \mathbf{e}_t, \quad (2.7)$$

donde  $\mathbf{f}(\mathbf{Y}_{t-1}) = (h(Y_{t-1}, \dots, Y_{t-p}), Y_{t-1}, \dots, Y_{t-p+1})'$ . Para comprobar las condiciones bajo las que el proceso  $Y_t$  es estacionario será necesario comprobar que  $Y_t$  es un proceso geoméricamente ergódico.

**Definición 2.1 (Proceso Geométricamente Ergódico).** El proceso  $Y_t$  es geoméricamente ergódico si existe una medida de probabilidad  $\pi$  definida sobre el espacio de los estados generado por  $(\mathbb{R}, \mathcal{B})$ , siendo  $\mathcal{B}$  la colección de conjuntos de Borel, un valor  $\rho$  constante positivo ( $\rho < 1$ ), y una función  $h$  medible no negativa y  $\pi$ -integrable tal que

$$\|\mathbf{P}^n(x, \cdot) - \pi(\cdot)\|_\tau \leq \rho^n h(x), \quad x \in \mathbb{R},$$

donde  $\|\cdot\|$  es una medida de la variación total.

Suponiendo que en el modelo (2.6)  $h(\cdot)$  es una función medible y que  $e_t$  tiene una función de densidad positiva con  $E[e_t] = 0$ . Entonces, sea  $\mathbf{x} = (x_1, \dots, x_p)'$ , el modelo inducido de Markov (2.7) es geoméricamente ergódico si alguna de las siguientes condiciones se mantiene:

C1 Existen los valores  $\rho \in (0, 1)$ ,  $c$  y  $a_i \geq 0$ , siendo  $\sum_{i=1}^p a_i = 1$ , para los cuales se cumple que

$$|h(\mathbf{x})| \leq \rho(a_1 |x_1| + \dots + a_p |x_p|) + c,$$

(Bhattacharya y Lee, 1995).

C2 La función  $h(\cdot)$  está acotada sobre conjuntos acotados y se cumple que

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} \frac{|h(\mathbf{x}) - (b_1 x_1 + \dots + b_p x_p)|}{\|\mathbf{x}\|} = 0,$$



donde  $\|\mathbf{x}\| = (x_1^2 + \dots + x_p^2)^{1/2}$ , y los valores  $b_1, \dots, b_p$  son constantes que satisfacen que  $1 - b_1 z - \dots - b_p z^p \neq 0$ , para todos los valores de  $z$  tales que  $|z| \leq 1$  (An y Huang, 1996).

C3 Existen los valores  $\delta \in (0, 1)$  y  $c$  para los cuales se cumple que

$$|h(\mathbf{x})| \leq \delta \max\{|x_1|, \dots, |x_p|\} + c,$$

(An y Huang, 1996).

A partir de estas condiciones, el proceso  $Y_t$  que sigue un modelo SETAR (2.2) es geoméricamente ergódico si se cumple alguna de las siguientes condiciones:

- i)  $\max_{1 \leq i \leq k} \sum_{j=1}^{p_i} |\phi_j^{(i)}| < 1$ , lo que corresponde a la condición C1 de ergodicidad.
- ii)  $\max_{1 \leq i \leq k} |\phi_j^{(i)}| < a_j$  y  $a_1 + \dots + a_p = 1$ , donde  $p = \max_{1 \leq i \leq k} p_i$ , lo que corresponde a la condición C3 de ergodicidad.

Estas condiciones para la ergodicidad son suficientes pero no necesarias.

**Teorema 2.2 (Estacionariedad, Fan y Yao(2005)).** *Suponiendo que el modelo de Markov (2.7) es geoméricamente ergódico. Entonces, existe una distribución estacionaria  $F$  tal que la serie temporal  $\{Y_t, t = 1, 2, \dots\}$  definida por (2.6) e inicializada según  $(X_0, X_{-1}, \dots, X_{-p+1})^T \sim F$  es estacionaria en sentido estricto.*

Entonces, si el proceso  $e_t$  en (2.2) es idénticamente distribuido para todos los regímenes, lo que implica que  $\sigma_{(1)} = \dots = \sigma_{(k)}$ , el modelo (2.7) será otra forma de escribir el modelo (2.2). Así pues, cuando  $Y_t$  sea geoméricamente ergódico será también estacionario.

## 2.3 Estimación de modelos SETAR

En esta sección se asume que en el modelo SETAR (2.2) los valores de  $k$ ,  $p_i$  y  $\Omega_i$  son conocidos para todo  $i = 1, \dots, k$ . Es decir, solo se necesitan estimar los parámetros  $\phi_j^{(i)}$  y las desviaciones  $\sigma_{(i)}$  en cada régimen.

### 2.3.1 Mínimos cuadrados ordinarios

El método de mínimos cuadrados ordinarios (OLS, *Ordinary Least Squares*) fue introducido por Legendre (1805) y Gauss (1809). En realidad, Gauss ideó el método en 1795 con 18 años de edad. La prueba de que descubrió el método años antes de publicarlo es la localización del asteroide Ceres. Dicho asteroide fue descubierto en 1801 siendo su

trayectoria seguida hasta que se ocultó tras el Sol. Con los datos obtenidos durante ese seguimiento, se intentó predecir la localización por la que Ceres aparecería de nuevo. De todos los intentos que se hicieron, la predicción de Gauss empleando el análisis de mínimos cuadrados fue el único que permitió su localización.

Los estimadores por mínimos cuadrados de los coeficientes autorregresivos  $\phi^{(i)} = (\phi_0^{(i)}, \dots, \phi_{p_i}^{(i)})$ , con  $i = 1, \dots, k$ , serán aquellos  $\{\hat{\phi}^{(1)}, \dots, \hat{\phi}^{(k)}\} \in \mathbb{R}$  que minimicen la función de costes

$$\sum_{i=1}^k S(\phi^{(i)}, d; \Omega_i), \quad (2.8)$$

siendo la función de costes en cada régimen

$$S(\phi^{(i)}, d; \Omega_i) = \sum_{\substack{Y_{t-d} \in \Omega_i \\ h < t \leq n}} \left\{ Y_t - (\phi_0^{(i)} + \phi_1^{(i)} Y_{t-1} + \dots + \phi_{p_i}^{(i)} Y_{t-p_i}) \right\}^2, \quad h = \max(p_i, d). \quad (2.9)$$

La estimación del modelo SETAR (2.2) se llevará a cabo mediante una proceso de dos etapas. Primero se minimiza la expresión (2.9) para cada valor del parámetro *delay*  $d$ . Dicho valor, se estima a continuación, siendo  $\hat{d}$  aquel que minimice (2.8). La minimización de la expresión (2.9) es equivalente a calcular los estimadores  $\hat{\phi}^{(1)}, \dots, \hat{\phi}^{(k)}$  mediante el método de OLS. Para hallar la forma de estos estimadores se reescribe el modelo (2.2) en forma matricial,

$$\mathbf{Y}_t = \sum_{i=1}^k (\mathbf{X}_t \phi^{(i)} + \mathbf{e}_t) \circ \mathbf{I}(\mathbf{Y}_{t-d} \in \Omega_i), \quad (2.10)$$

donde el vector  $\mathbf{Y}_t = (Y_h, \dots, Y_n)'$ , siendo  $h = \max(1, p + 1 - d)$ . La matriz de variables explicativas es  $\mathbf{X}_t = (\mathbf{1}_n, \mathbf{Y}_{t-1}, \dots, \mathbf{Y}_{t-p_i})$ , siendo  $\mathbf{1}_n$  un vector de 1 de dimensión  $n$ , y el vector de parámetros a estimar es  $\phi^{(i)} = (\phi_0^{(i)}, \dots, \phi_{p_i}^{(i)})$ . El vector  $\mathbf{e}_t$  contiene los  $k$  procesos  $\{e_t^{(i)}\}$ . El vector  $\mathbf{I}(\cdot)$  tomará valores 1 cuando se cumpla la condición y 0 en los otros casos. Por último, notar que el operador  $\circ$  simboliza el producto de Hadamard, es decir, la multiplicación matricial elemento a elemento.

Así pues, el estimador OLS para cada uno de los regímenes del modelo (2.10) será

$$\hat{\phi}^{(i)} = [\mathbf{X}_t^{(i)'} \mathbf{X}_t^{(i)}]^{-1} \mathbf{X}_t^{(i)'} \mathbf{Y}_t^{(i)}, \quad i = 1, \dots, k, \quad (2.11)$$

donde  $\mathbf{X}_t^{(i)}$  y  $\mathbf{Y}_t^{(i)}$  son los valores correspondientes a las observaciones pertenecientes al régimen  $i$ , y se calculan según  $\mathbf{X}_t^{(i)} = \mathbf{X}_t \circ [\mathbf{I}(\mathbf{Y}_{t-d} \in \Omega_i) \mathbf{1}_{p_i}']$  y  $\mathbf{Y}_t^{(i)} = \mathbf{Y}_t \circ \mathbf{I}(\mathbf{Y}_{t-d} \in \Omega_i)$ ,

respectivamente. Un estimador para la desviación  $\sigma_{(i)}$  en cada régimen es

$$\hat{\sigma}_{(i)} = \frac{1}{n_i} S \left( \hat{\phi}^{(i)}, \hat{d}_i; \Omega_i \right), \quad (2.12)$$

donde  $n_i$  es el número de elementos en cada una de las particiones del espacio de los estados.

### 2.3.2 Propiedades del estimador OLS

Las propiedades asintóticas del estimador dependerán de la relación entre el proceso  $Y_t$  y sus retardos, es decir, dependerá de la función de regresión  $E[Y_t|Y_{t-k} = y_{t-k}, k \geq 1]$ . Para ilustrar las diferentes formas que puede tomar dicha función de regresión, se generaron diferentes realizaciones provenientes del modelo (2.2) con valores  $k = 2, p_1 = p_2 = 1, \phi_1^{(1)} = .6$  y  $\phi_1^{(2)} = -.6$ . Emplear diferentes valores del umbral  $r$  nos permitirá ver las dos formas posibles de la función de regresión  $E[Y_t|Y_{t-1}]$ . Por una parte en la figura (2.1a) se ve como la función de regresión es continua, puesto que el valor de los parámetros evoluciona de manera continua a lo largo del dominio, mientras que en (2.1b) vemos como dicha evolución es discontinua a la altura del valor del umbral  $r$ .

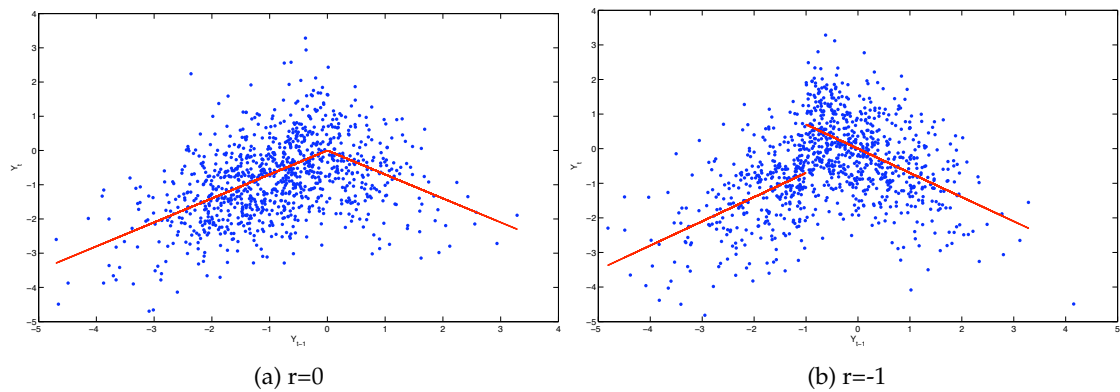


Figura 2.1: Gráfico de dispersión de  $Y_t$  respecto a la variable umbral  $Y_{t-1}$

Para el caso discontinuo, Chan (1993) mostró que, bajo ciertas condiciones, los estimadores de los parámetros del modelo son consistentes en sentido estricto, como podemos ver en el teorema 2.3.

**Teorema 2.3 (Chan, 1993).** Sea  $\{Y_t\}$  un proceso generado por  $\theta = (\phi^{(1)}, \phi^{(2)}, r, d)$  y  $(\sigma^{(1)}, \sigma^{(2)})$ , que satisface el modelo (2.2), con  $k = 2$  y  $p_1 = p_2 = p$ . Suponiendo que  $\{Y_t\}$  es geométricamente ergódico y estacionario con segundo momento finito y que la distribución conjunta de

$(Y_1, \dots, Y_p)$  admite una densidad positiva. Entonces, el estimador  $\hat{\theta}$  es consistente, es decir,

$$\hat{\theta} \xrightarrow{c.s.} \theta.$$

y los estimadores  $(\hat{\sigma}^{(1)}, \hat{\sigma}^{(2)})$  también lo son.

Además del resultado de consistencia para los estimadores, Chan obtuvo los siguientes resultados acerca de las distribuciones límite de dichos estimadores.

**Teorema 2.4 (Chan, 1993).** Suponiendo que las condiciones del teorema 2.3 se mantienen. Suponiendo, además, que:

- i) La cadena de Markov es  $\mathbf{Y}_t = (Y_t, Y_{t-1}, \dots, Y_{t-p+1})'$  es geométricamente ergódica, y  $E[\mathbf{Y}_t^4] < \infty$ .
- ii)  $\{e_t\}$  tiene una función de densidad continua y positiva, y  $E[e_t^4] < \infty$ .
- iii) La función autorregresiva es discontinua, es decir, existe  $\mathbf{z} = (1, z_{p-1}, z_{p-2}, \dots, z_0)'$  tal que  $(\phi^{(1)} - \phi^{(2)})\mathbf{z} \neq 0$  y  $z_{p-d} = r$ .

Entonces  $(\hat{r} - r)$  es asintóticamente independiente de  $(\hat{\phi}^{(1)} - \phi^{(1)}, \hat{\phi}^{(2)} - \phi^{(2)})$  y  $n(\hat{r} - r) \equiv O_p(1)$ . Además  $\sqrt{n_i}(\hat{\phi}^{(i)} - \phi^{(i)})$  es asintóticamente normal

$$\sqrt{n_i}(\hat{\phi}^{(i)} - \phi^{(i)}) \xrightarrow{d} N(0, \sigma^{(i)} \mathbf{W}_i^{-1/2}), \quad i = 1, 2;$$

donde

$$\mathbf{W}_i = \begin{pmatrix} 1 & \mu \mathbf{1}_p' \\ \mu \mathbf{1}_p & E(\boldsymbol{\xi}_t^{(i)} \boldsymbol{\xi}_t^{(i)'} ) \end{pmatrix}, \quad \boldsymbol{\xi}_t^{(i)} = (\xi_1^{(i)}, \dots, \xi_p^{(i)}),$$

siendo  $\mu = E[\boldsymbol{\xi}_t]$  y

$$\boldsymbol{\xi}_t^{(i)} = \phi_0^{(i)} + \phi_1^{(i)} \boldsymbol{\xi}_{t-1}^{(i)} + \dots + \phi_p^{(i)} \boldsymbol{\xi}_{t-p}^{(i)} + \varepsilon_t.$$

El teorema 2.4 nos muestra que la tasa de convergencia de  $\hat{r}$  es del orden de  $n^{-1}$ , mientras que los parámetros autorregresivos  $\hat{\phi}^{(i)}$  son  $n^{-1/2}$ .

La forma de  $E[Y_t | Y_{t-k} = y_{t-k}, k \geq 1]$  queda dispuesta por la condición iii) del teorema 2.4. Entonces, si el proceso sigue un modelo SETAR con 2 regímenes, la función de regresión será continua si se cumple que  $(\phi^{(1)} - \phi^{(2)})\mathbf{z} = 0$ , o lo que es lo mismo, se cumple que  $\phi_j^{(1)} = \phi_j^{(2)}$ , para  $1 \leq j \neq d \leq p$ , y  $\phi_0^{(1)} + r\phi_d^{(1)} = \phi_0^{(2)} + r\phi_d^{(2)}$ . Así pues, el modelo (2.2), con  $k = 2$  y  $p_1 = p_2 = p$  se podrá escribir como

$$Y_t = \phi_0^{(1)} + r\phi_d^{(1)} + \sum_{j=1; j \neq d}^p \phi_j^{(1)} Y_{t-j} + \begin{cases} \phi_d^{(1)} (Y_{t-d} - r) + e_t^{(1)} & , \text{ si } Y_{t-d} - r \leq 0 \\ \phi_d^{(2)} (Y_{t-d} - r) + e_t^{(2)} & , \text{ si } Y_{t-d} - r > 0 \end{cases} \quad (2.13)$$

Chan y Tsay (1998) mostraron resultados similares a los obtenidos por Chan (1993) en los teoremas 2.3 y 2.4 para el caso continuo. Por una parte el teorema 2.3 en el caso continuo es semejante al obtenido en el caso discontinuo, con la salvedad de la necesidad de incluir una condición extra,  $\phi_d^{(1)} \neq \phi_d^{(2)}$ , que garantiza la identificabilidad del valor del umbral.

**Teorema 2.5 (Chan y Tsay, 1998).** Sea  $\{Y_t\}$  un proceso generado por  $\theta = (\phi^{(1)}, \phi^{(2)}, r, d)$  y  $(\sigma^{(1)}, \sigma^{(2)})$ , que satisface el modelo (2.13), con  $k = 2$  y  $p_1 = p_2 = p$ . Suponiendo que  $\{Y_t\}$  es geométricamente ergódico y estacionario con segundo momento finito, que la distribución conjunta de  $(Y_1, \dots, Y_p)$  admite una densidad positiva y que  $\phi_d^{(1)} \neq \phi_d^{(2)}$ . Entonces, el estimador  $\hat{\theta}$  es consistente, es decir,

$$\hat{\theta} \xrightarrow{c.s.} \theta.$$

y los estimadores  $(\hat{\sigma}^{(1)}, \hat{\sigma}^{(2)})$  también lo son.

Para obtener los resultados de la distribución límite, antes será necesario definir las condiciones que requiere el proceso  $Y_t$  para ser  $\beta$ -mixing (para ampliar la información sobre este tipo de procesos ver, p.ej., Doukhan, 1994).

**Definición 2.6 (Coeficiente  $\beta$ -mixing).** Sean  $\Xi$  y  $\Upsilon$  dos  $\sigma$ -álgebras, entonces una medida de su dependencia es su coeficiente  $\beta$ -mixing

$$\beta(\Xi, \Upsilon) = \frac{1}{2} \sup \sum_{(i,j) \in I \times J} |P(\Xi_i \cap \Upsilon_j) - P(\Xi_i)P(\Upsilon_j)|,$$

donde el supremo es tomado sobre todas las particiones finitas  $(\Xi_i)_{i \in I}$  y  $(\Upsilon_j)_{j \in J}$  de los espacio muestrales medibles  $\Xi$  y  $\Upsilon$ , respectivamente.

**Definición 2.7 (Proceso  $\beta$ -mixing).** El proceso  $\{Y_t\}$  es  $\beta$ -mixing con tasa geométrica si

$$\beta\{\sigma(\dots, Y_{-1}, Y_0), \sigma(Y_t, Y_{t+1}, \dots)\} = O(\rho^t),$$

para algún  $0 \leq \rho < 1$  y  $\sigma(\cdot)$  denota la  $\sigma$ -álgebra generada por las variables aleatorias que contienen.

Así pues, de las definiciones de un proceso geométricamente ergódico 2.1 y de un proceso  $\beta$ -mixing 2.7 se deduce que una cadena de Markov estacionaria y geométricamente ergódica es  $\beta$ -mixing con una tasa que decrece geométricamente. Una vez definidos los conceptos necesarios se muestra el resultado obtenido por Chan y Tsay (1998).

**Teorema 2.8 (Chan y Tsay, 1998).** Sea  $\{Y_t\}$  un proceso generado por  $\theta = (\phi^{(1)}, \phi^{(2)}, r, d)$  y  $(\sigma^{(1)}, \sigma^{(2)})$ , que satisface el modelo (2.13), con  $k = 2$  y  $p_1 = p_2 = p$ . Suponiendo que  $\{Y_t\}$  es

un proceso estacionario  $\beta$ -mixing con una tasa que decrece geométricamente, y que la densidad estacionaria de  $Y_t$  es positiva y acotada entorno al verdadero valor del umbral  $r$ . Suponemos además que  $\phi_d^{(1)} \neq \phi_d^{(2)}$  y que, para algún  $q > 2$ ,  $E(|Y_t|^q) < \infty$ . Entonces, sean  $U = E(H_t H_t')$  y  $V = E(\hat{e}_t^2 H_t H_t')$  la distribución asintótica es

$$\sqrt{n} (\hat{\theta} - \theta) \xrightarrow{d} N(0, U^{-1} V U^{-1}),$$

donde  $H_t$  denota la derivada parcial de  $\hat{e}_t$  con respecto a  $\theta$ .

## 2.4 Selección de la estructura autorregresiva

Los criterios de selección de modelos son empleados para encontrar el orden óptimo de un modelo paramétrico basándose en los datos observados, siendo el orden óptimo aquel que minimice el criterio de selección empleado. En esta sección se revisan las propuestas existentes para selección de modelos SETAR. Para ver una amplia comparativa de múltiples criterios de selección existentes para modelos por umbrales puede verse, por ejemplo, Kapetanios (2001).

### 2.4.1 Criterio de información de Akaike

El criterio de información de Akaike (AIC, *Akaike Information Criteria*), propuesto por Akaike (1973), se basa en la idea de minimizar la diferencia entre el modelo ajustado a los datos y el verdadero. Para ello, Akaike propuso emplear la divergencia de Kullback-Leibler que da una medida de la diferencia entre las densidades  $f$  y  $g$ , y que viene definida por

$$KL(g||f) = \int g(x) \log[g(x)] dx - \int g(x) \log[f(x)] dx, \quad (2.14)$$

Reescribiendo (2.14) y empleando la desigualdad de Jensen vemos que

$$\begin{aligned} KL(g||f) &= -E \left\{ \log \left[ \frac{f(x)}{g(x)} \right] \right\} \geq -\log \left\{ E \left[ \frac{f(x)}{g(x)} \right] \right\} \\ &= -\log \left( \int g(x) \frac{f(x)}{g(x)} dx \right) = 0, \end{aligned}$$

donde la igualdad se mantendrá si y sólo si  $f = g$ . Así pues, una buena aproximación de las densidades  $f$  y  $g$  será conseguir que la medida (2.14) sea lo menor posible. De esta manera se justifica que la elección del modelo venga dado por el valor mínimo del criterio de selección empleado.

Akaike (1973) desarrolla a partir de la divergencia de Kullback-Leibler el criterio de selección AIC, cuya forma para modelos SETAR será,

$$\text{AIC}([p_1, \dots, p_k]) = \sum_{i=1}^k \left[ n_i \log \left\{ \hat{\sigma}_{(i)}^2 \right\} + 2(p_i + 1) \right]. \quad (2.15)$$

Tong (1983, 1990) propuso seleccionar los ordenes autorregresivos de los diferentes regímenes en un modelo SETAR empleando el criterio AIC, aunque no mostró ninguna justificación teórica.

Hurvich y Tsai (1989) propusieron una modificación del AIC para obtener una mejor corrección del sesgo, el criterio de información de Akaike corregido (AICc, *corrected Akaike Information Criteria*). El AICc para modelos SETAR fue hallado por Wong y Li (1998) y está definido como

$$\text{AICc}([p_1, \dots, p_k]) = \sum_{i=1}^k \left[ n_i \log \left\{ \hat{\sigma}_{(i)}^2 \right\} + \frac{n_i(n_i + p_i + 1)}{(n_i - p_i - 3)} \right]. \quad (2.16)$$

McQuarrie et al. (1997) muestran que la corrección del sesgo del AICc no es adecuada. Por ello propusieron un criterio de información asintóticamente insesgado. La forma del AICu (*unbiased Akaike Information Criteria*) para modelos SETAR, deducida por De Gooijer (2001) es

$$\text{AICu}([p_1, \dots, p_k]) = \text{AICc}([p_1, \dots, p_k]) + \sum_{i=1}^k \left[ n_i \log \left\{ \frac{n_i}{n_i - p_i - 2} \right\} \right]. \quad (2.17)$$

### 2.4.2 Criterios de información Bayesianos

El principal problema de los criterios de información basados en la idea de Akaike es que la selección de los ordenes autorregresivos no es consistente. Una de las propuestas más populares que soluciona el problema de la falta de consistencia es el criterio de información Bayesiano (BIC, *Bayesian Information Criteria*). Su nombre se debe a que el criterio fue desarrollado empleando argumentos Bayesianos (ver, p.ej., Akaike, 1977; Schwarz, 1978). La forma del criterio BIC para modelos SETAR será

$$\text{BIC}([p_1, \dots, p_k]) = \sum_{i=1}^k \left[ n_i \log \left\{ \hat{\sigma}_{(i)}^2 \right\} + (p_i + 1) \log(n_i) \right]. \quad (2.18)$$

Más recientemente, se han propuesto técnicas bayesianas de selección de modelos basadas en métodos de cadenas de Markov Monte Carlo (p.ej., Campbell, 2004; Unnikr-

ishnan, 2004).

### 2.4.3 Criterios basados en Validación Cruzada

El concepto de validación cruzada se basa en la división de la muestra en dos submuestras disjuntas, una de las cuales es la de calibración que se emplea para estimar un modelo y otra la de validación que se emplea para evaluar el rendimiento de dicho modelo. Para encontrar más información sobre el tema ver, por ejemplo, Stone (1974).

Stoica et al. (1986) proponen la siguiente generalización de esta idea. Sea  $m$  un valor positivo entero y  $n$  el tamaño muestra, se define el entero  $L$  que vendrá definido por  $L = \lfloor n/m \rfloor$ , donde  $\lfloor \cdot \rfloor$  denota la parte entera. Entonces, la idea es dividir el proceso en submuestras, de manera que se creen  $L$  submuestras de calibración y  $L$  submuestras de validación. Siguiendo esta idea Stoica et al. (1986) proponen un criterio de selección para modelos lineales. Sea  $\hat{\phi}_i$  la estimación de un modelo AR(p) en la submuestra de calibración  $i$ -ésima y sea  $I_i$  la muestra de validación  $i$ -ésima, entonces el criterio de selección basado en validación cruzada se calcula mediante

$$CV_m(p) = \sum_{i=1}^L \sum_{t \in I_i} \left\{ Y_t - X_t \hat{\phi}_i \right\}^2, \quad (2.19)$$

Como puede verse el criterio (2.19) dependerá de la elección de  $m$ . Stone (1974) recomienda emplear  $m = 1$ , en cuyo caso nos encontramos ante el método habitualmente conocido como validación cruzada "dejar-uno-fuera" ("leave-one-out"). El problema es que la propuesta de Stone (1974) fue en el ámbito de modelos de regresión lineal, no para el ámbito de las series temporales, a pesar de ello De Gooijer (2001) emplea  $m = 1$ .

El criterio de selección basado en validación cruzada para modelos SETAR fue propuesto por De Gooijer (2001), y viene definido por

$$C_1([p_1, \dots, p_k]) = \sum_{i=h}^L \sum_{t=i-m+1}^{im} \hat{e}_t^2 \left\{ \hat{\theta}_i([p_1, \dots, p_k]) \right\}, \quad (2.20)$$

donde  $\hat{\theta}_i$  es la estimación en la submuestra de validación  $i$ -ésima de los parámetros autorregresivos del modelo SETAR (2.2) y los residuos predictivos  $\hat{e}_t^2$  serán calculados

$$\hat{e}_t^2 \left\{ \hat{\theta}_i([p_1, \dots, p_k]) \right\} = Y_{t,i} - \hat{Y}_{t,i}, \quad (2.21)$$

donde  $\hat{Y}_{t,i}$  es la predicción un paso adelante empleando la muestra de calibración  $i$ -ésima de la observación  $Y_{t,i}$  de la muestra  $i$ -ésima de validación. De Gooijer (2001) siguiendo el espíritu de las modificaciones que se han ido efectuando en el criterio AIC,



propuso un criterio de validación cruzada corregido definido como

$$C_C([p_1, \dots, p_k]) = n \log \left\{ \frac{1}{n} C_1([p_1, \dots, p_k]) \right\} + \sum_{j=1}^k \frac{2(p_j + 2)(p_j + 3)}{n_j - p_j - 3}. \quad (2.22)$$

De igual manera propuso un criterio de validación cruzada insesgado, que puede ser definido según

$$C_U([p_1, \dots, p_k]) = n \log \left\{ \frac{1}{n} C_1([p_1, \dots, p_k]) \right\} + \sum_{j=1}^k \left\{ n_j \log \left( \frac{n_j}{n_j - p_j - 2} \right) \frac{2(p_j + 2)(p_j + 3)}{n_j - p_j - 3} \right\}. \quad (2.23)$$

#### 2.4.4 Criterios de selección Bootstrap

Öhrvik y Schoier (2005) proponen un procedimiento basado en métodos Bootstrap para seleccionar los órdenes de modelos SETAR. Estos criterios se basan en la estimación del error de predicción por una media ponderada de la tasa de error aparente de la muestra y de la tasa de error media obtenida de muestras Bootstrap que no contienen a la observación que estamos prediciendo. De esta manera corrige la subestimación del error producido por la estimación del error aparente. Sea  $\hat{Y}_{t,g}$  la predicción de  $Y_t$  empleando el  $g$ -ésimo modelo, el error aparente vendrá dado por

$$\hat{\varepsilon}_g = \frac{1}{n-p} \sum_{t=p+1}^n \left( Y_t - \hat{Y}_{t,g} \right)^2. \quad (2.24)$$

El problema de los métodos Bootstrap es que asumen que el remuestreo con reemplazamiento se realiza sobre observaciones independientes e idénticamente distribuidas, asunción que habitualmente no se cumple en series temporales. Para solucionarlo Öhrvik y Schoier (2005) emplean diferentes aproximaciones para obtener las muestras Bootstrap. Una vez obtenidas las  $B$  muestras Bootstrap, se puede diferenciar entre aquellas en las cuales la observación a predecir  $Y_t$  está incluida y en las que no lo está. La idea de Öhrvik y Schoier (2005) es emplear la media del error de las muestras Bootstrap que no contienen al punto que está siendo predicho para ajustar la tasa de error aparente. Dicha media vendrá definida por

$$\hat{\varepsilon}_g^B = \frac{1}{n-p} \sum_{t=p+1}^n \sum_{b \in A_t} \frac{\left( Y_t - \hat{Y}_{t,g}^b \right)^2}{\#A_t}, \quad (2.25)$$

donde  $A_t$  son los índices de las muestras Bootstrap que no contienen la  $t$ -ésima ob-

servación e  $\hat{Y}_{t,g}^b$  es la predicción de  $Y_t$  empleando el  $g$ -ésimo modelo sobre la  $b$ -ésima muestra Bootstrap. Entonces, el criterio Bootstrap para seleccionar modelos SETAR viene definido por

$$\text{BSC}([p_1, \dots, p_k]) = \hat{\varepsilon}_g + (1 - \exp(-1)) [\hat{\varepsilon}_g^B - \hat{\varepsilon}_g]. \quad (2.26)$$

Öhrvik y Schoier (2005) proponen además modificaciones del criterio corrigiendo los mismos defectos que las correcciones efectuadas al AIC. Entonces el criterio Bootstrap corregido será

$$\text{BSCc}([p_1, \dots, p_k]) = (n - p) \log \{ \text{BSC}([p_1, \dots, p_k]) \} - 2 \sum_{j=1}^k (p_j + 1) + \sum_{j=1}^k \frac{n_j (n_j + p_j + 1)}{n_j + p_j - 3}, \quad (2.27)$$

y el criterio Bootstrap insesgado vendrá definido por

$$\text{BSCu}([p_1, \dots, p_k]) = \text{BSC}([p_1, \dots, p_k]) + \sum_{j=1}^k n_j \log \left\{ \frac{n_j}{n_j - p_j - 2} \right\}. \quad (2.28)$$

## 2.4.5 Mejora de criterios de selección

Hurvich et al. (1990) proponen una mejora del criterio AIC para su uso en selección de ordenes de un proceso autorregresivo en muestras pequeñas. Su mejora se basa en una aproximación de la esperanza de la divergencia de Kullback-Leibler basada en el determinante de la estimación de la matriz de covarianzas del proceso autorregresivo.

Dado que el modelo SETAR se compone de procesos autorregresivos, Galeano y Peña (2007) emplean la idea de Hurvich para mejorar varios de los criterios de selección anteriormente mostrados. La modificación de los criterios consiste en la adición del determinante de la matriz de covarianzas al criterio de selección que se va a modificar. Por ejemplo, en el caso del AIC, la modificación propuesta por Galeano y Peña (2007), que denotaron como AIC\*, es

$$\text{AIC}^*([p_1, \dots, p_k]) = \sum_{i=1}^k \left[ n_i \log \left\{ \hat{\sigma}_{(i)}^2 \right\} + 2(p_i + 1) + \log \left| \mathbf{Q} \left( \hat{\phi}^{(i)} \right) \right| \right], \quad (2.29)$$

donde el determinante en cada régimen se calcula empleando (p.ej., van der Leeuw, 1994)

$$\left| \mathbf{Q} \left( \hat{\phi}^{(i)} \right) \right| = \frac{1}{|\mathbf{M}'\mathbf{M} - \mathbf{N}\mathbf{N}'|}, \quad (2.30)$$

siendo  $\mathbf{M}$  y  $\mathbf{N}$  matrices de dimensión  $(p_i \times p_i)$  que contienen los elementos dados por

$$\mathbf{M}_{ab} = \begin{cases} 0 & , \text{ si } a < b \\ 1 & , \text{ si } a = b \\ -\hat{\phi}_{a-b}^{(i)} & , \text{ si } a > b \end{cases}, \quad \mathbf{N}_{ab} = \begin{cases} -\hat{\phi}_{p_i+(a-b)}^{(i)} & , \text{ si } a \leq b \\ 0 & , \text{ si } a > b \end{cases}. \quad (2.31)$$

De manera análoga a la mostrada para el criterio AIC, Galeano y Peña (2007) proponen modificar varios criterios de selección como son el AICc, el BIC, el  $C_1$ , el Cc y el Cu.

## 2.5 Identificación de modelos SETAR

La realización de un contraste de linealidad es una buena forma de comprobar si el modelo SETAR empleado resume de manera adecuada la información contenida por los datos. Una de las complicaciones de los contrastes para modelos SETAR es que los umbrales  $r_i, i = 1, \dots, k - 1$ , no están identificados en la hipótesis nula, por lo que la distribución asintótica habitual del estadístico no puede ser aplicada. Así pues será necesario realizar otra aproximación de la distribución.

Además, el uso de un modelo por umbrales conlleva la necesidad de estimar cuál es la variable umbral y la partición del espacio de los estados. La estimación de la variable umbral no es una gran complicación puesto que se pueden probar diferentes variables y seleccionar entre ellas empleando algún criterio de selección.

El problema más complicado es la identificación de la partición del espacio de los estados, o lo que es lo mismo la estimación de los umbrales. Tsay (1989) y Tong (1990) identifican los valores de los umbrales mediante un análisis visual de diferentes gráficos. Otra opción propuesta es emplear algoritmos de búsqueda intensiva probando diferentes valores posibles.

### 2.5.1 Contrastes de linealidad

Sea  $Y_t$  un proceso estrictamente estacionario, se quiere contrastar la hipótesis nula

$$H_0 : Y_t = \phi_0 + \sum_{j=1}^p \phi_j Y_{t-j} + e_t, \quad (2.32)$$

contra la alternativa

$$H_1 : Y_t = \phi_0 + \sum_{j=1}^p \phi_j Y_{t-j} + \left[ \delta_0 + \sum_{j=1}^p \delta_j Y_{t-j} \right] I(Y_{t-d} \leq r) + e_t. \quad (2.33)$$

El contraste a emplear será diferente dependiendo de la forma del proceso  $\{e_t\}$ , es decir, de si es homocedástico o heterocedástico. En ambos casos se supone que el orden del proceso autorregresivo  $p$  y el valor del parametro *delay*  $d$  son conocidos y que el valor del umbral  $r$  permanece dentro de un intervalo conocido  $I_r$ . Este intervalo es tomado, en general, seleccionando un porcentaje de puntos en el centro de la muestra. En esta sección se repasan los diferentes contrastes de linealidad que se han ido proponiendo, centrándonos en aquellos contrastes específicos en los que la hipótesis alternativa a la linealidad es la existencia de un modelo SETAR.

### Contraste de ratio de verosimilitud

Sea  $\{e_t\}$  un proceso independiente e idénticamente distribuido según  $N(0, \sigma)$ , siendo  $\sigma < \infty$ . Entonces, dado que el proceso  $\{e_t\}$  es gaussiano y homocedástico, se emplea un contraste de la F. Así pues, la hipótesis nula (2.32) se rechaza para valores grandes del estadístico

$$F_n = T \frac{(\hat{\sigma}^2 - \hat{\sigma}_0^2)}{\hat{\sigma}^2}, \quad (2.34)$$

donde  $T = \{n - \max(d, p)\}$  y las varianzas  $\hat{\sigma}_0^2$  y  $\hat{\sigma}^2$  son las estimaciones OLS empleando (2.12) de los modelos (2.32) y (2.33), respectivamente. El problema, como dijimos anteriormente, es que la distribución asintótica del estadístico  $F_n$  no es una ji-cuadrado de Pearson ya que el parámetro desconocido  $r$  está ausente en la hipótesis nula (2.32), por lo que una nueva aproximación se hace necesaria.

La aproximación propuesta en los trabajos de Chan y Tong (1990), Chan (1990) y Chan (1991), se basa en la familia de procesos de Poisson (ver, Aldous 1989). Según la propuesta de los autores, los niveles de significación del estadístico (2.34) se aproximan mediante

$$P\{F_n > x | H_0\} \approx 1 - \exp \left\{ -2\chi_{p+1}^2(x) \left( \frac{x}{p+1} - 1 \right) \sum_{i=1}^{p+1} \int_{I_r} h_i(y) dy \right\}, \quad (2.35)$$

donde  $\chi_j^2(\cdot)$  denota la función de densidad de probabilidad de la distribución ji-cuadrado de Pearson con  $j$  grados de libertad y

$$h_i(y) = \frac{dJ_i(y)}{dy}, \quad (2.36)$$

siendo

$$J_i(y) = \frac{1}{2} \log \left\{ \frac{P_0(Y_t \leq y)}{P_0(Y_t > y)} \right\}, \quad 1 \leq i < p, \quad (2.37)$$

donde  $J_p(y)$  y  $J_{p+1}(y)$  son las raíces de la ecuación  $x^2 - bx + c = 0$  con

$$b = E_0 \left\{ \left( 1 + \frac{Y_t^2}{E_0(Y_t^2)} \right) I(Y_t \leq y) \right\}; \quad (2.38)$$

$$c = \frac{1}{E_0(Y_t^2)} \left[ P_0(Y_t \leq y) E_0 \{ Y_t^2 I(Y_t \leq y) \} - \{ E_0 \{ Y_t I(Y_t \leq y) \} \}^2 \right]. \quad (2.39)$$

Notar que en las expresiones anteriores  $P_0$  y  $E_0$  denotan, respectivamente, la probabilidad y la esperanza bajo la hipótesis nula  $H_0$ . En Chan (1991) se pueden encontrar tablas para diferentes ordenes  $p$  y para diferentes intervalos  $I_r$  compuestos por porcentajes de la muestra.

### Contrastes propuestos por Hansen

Hansen (1996, 1999) ha propuesto varios contrastes para modelos SETAR. En este trabajo nos hemos centrado en los que propuestos en Hansen (1996), que emplean las hipótesis (2.32) y (2.33). En Hansen (1999) propone un ampliación para permitir contrastar

$$\begin{cases} H_0 : AR(p) \\ H_1 : SETAR(2; p, d) \end{cases} \quad \text{y} \quad \begin{cases} H_0 : SETAR(2; p, d) \\ H_1 : SETAR(3; p, d) \end{cases} \quad (2.40)$$

Las diferentes propuestas existentes en Hansen (1996) se diferencian en la forma de aproximar la distribución asintótica del estadístico.

**Distribución asintótica basada en el contraste de Davies** Davies (1987) propuso un contraste cuando en la hipótesis alternativa existe un parámetro sin identificar en la hipótesis nula. Entonces, sea el parámetro sin identificar  $r^* \in I_r$ , el estadístico (2.34) se calcula según

$$F_n = \sup_{r^* \in I_r} (F_n(r^*)). \quad (2.41)$$

La idea del contraste es sencilla. Como explicamos anteriormente la distribución asintótica de  $F_n$  no será la distribución ji-cuadrado  $\chi^2(\cdot)$  habitual, ya que el valor de  $r$  no es conocido. Pero para un valor del umbral  $r^*$  fijado, la distribución de  $F_n(r^*)$  si será  $\chi^2(\cdot)$ . Entonces, la distribución del  $F_n$  sería el máximo de las diferentes ji-cuadrado. El problema es que hallar esto no es trivial, puesto que  $F_n(r^*)$  es una función aleatoria. Hansen (1996) muestra que la función aleatoria  $F_n(r^*)$  tiene una distribución asintótica Normal multivariante y que el estadístico  $F_n$  converge en distribución a  $S_n$ . La dificultad ahora se encuentra en que la distribución de  $S_n$  no puede ser caracterizada. Hansen (1996) muestra cómo calcular un valor de  $S_n$  proveniente de su distribución asintótica

mediante

$$S_n = \max_{r^* \in I_r} \hat{u}' M_n(r^*) \hat{u}, \quad (2.42)$$

donde

$$M_n(r^*) = \left[ \mathbf{X}_{r^*}^{1'} \mathbf{X}_{r^*}^1 - \mathbf{X}_{r^*}^{1'} \mathbf{X}_{r^*}^1 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_{r^*}^{1'} \mathbf{X}_{r^*}^1 \right]^{-1} \mathbf{X}_{r^*}^{1'}, \quad (2.43)$$

siendo  $\mathbf{X} = [\mathbf{Y}_h, \dots, \mathbf{Y}_n]'$ , donde  $\mathbf{Y}_t = (1, Y_{t-1}, \dots, Y_{t-p})'$  y  $\mathbf{X}_{r^*}^1 = \mathbf{X} \circ \mathbf{I}(\mathbf{Y}_{t-d} \leq r^*)$ . Además el vector  $\hat{u}$  se calcula empleando

$$\hat{u} = u - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' u, \quad (2.44)$$

donde  $u$  es un vector de números aleatorios con distribución  $N(0, \mathbf{I}_n)$ . Entonces, la distribución asintótica de  $S_n$  puede ser aproximada a partir de la realización de replicas provenientes de la expresión (2.42). De esta manera, ya que  $F_n$  converge en distribución a  $S_n$  se hallan los niveles de significación.

En el caso de que  $\{e_t\}$  sea un proceso heterocedástico, el desarrollo es idéntico al caso homocedástico, cambiando únicamente en el cálculo del vector de números aleatorios  $\hat{u}$ , que para el caso heterocedástico se hará mediante

$$\hat{u} = \eta - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \eta, \quad (2.45)$$

donde  $\eta = (u \circ e_t) / \sigma$ , siendo  $u$  un vector de números aleatorios con distribución  $N(0, \mathbf{I}_n)$ .

**Distribución asintótica basada en métodos Bootstrap** Hansen (1996) muestra que la distribución asintótica de (2.34) puede aproximarse mediante el siguiente procedimiento Bootstrap. Sea  $\{u_t\}$  un proceso de números aleatorios independientes muestreados a partir de una distribución  $N(0, 1)$ . Entonces, el estadístico Bootstrap se calcula mediante

$$F_n^B = \sup_{r^* \in I_r} (F_n(r^*)), \quad (2.46)$$

donde

$$F_n(r^*) = n \frac{\hat{\sigma}^2(r^*) - \hat{\sigma}_0^2}{\hat{\sigma}^2(r^*)}, \quad (2.47)$$

siendo  $\hat{\sigma}^2(r^*)$  la varianza residual de la regresión de los números aleatorios  $u_t$  sobre el proceso  $Y_t$  mostrado en (2.33). Hansen (1996) muestra que la distribución de  $F_n^B$  converge débilmente en probabilidad a la distribución de la hipótesis nula de  $F_n$  bajo diferentes alternativas para  $\delta$  en la expresión (2.33). Entonces, la aproximación Bootstrap a los niveles de significación asintóticos será el porcentaje de muestras Bootstrap en las cuales  $F_n^B$  es mayor que  $F_n$ .

El caso heterocedástico es diferente en el caso de la aproximación Bootstrap. Sea  $\{e_t\}$  un proceso independiente e idénticamente distribuido según  $N(0, \sigma_t)$ , siendo  $\sigma_t < \infty$ . En este caso, dado que el proceso  $\{e_t\}$  es heterocedástico habrá que emplear un contraste de Wald o uno de multiplicadores de Lagrange. Hansen (1996) propone un procedimiento para encontrar la distribución asintótica de un contraste de Wald, en el que el estadístico será

$$W_n = \sup_{r^* \in I_r} (W_n(r^*)), \quad (2.48)$$

siendo

$$W_n(r^*) = \theta(r^*)' R \left[ R M(r^*)^{-1} V(r^*) M(r^*)^{-1} R \right]^{-1} R' \theta(r^*), \quad (2.49)$$

donde  $\theta(r^*) = [\phi_0, \dots, \phi_p, \delta_0, \dots, \delta_p]'$  es el vector de parámetros del modelo (2.33), y la matriz seleccionadora  $R$  es una matriz diagonal de dimensión  $2(p+1) \times 2(p+1)$ , formada según

$$R_{ii} = \begin{cases} 1 & , \text{ si } i \leq p+1 \\ -1 & , \text{ si } i > p+1 \end{cases}, \quad (2.50)$$

además

$$M(r^*) = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_t(r^*) \mathbf{X}_t(r^*)', \quad (2.51)$$

y

$$V(r^*) = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_t(r^*) \mathbf{X}_t(r^*)' e_t^2, \quad (2.52)$$

siendo  $\mathbf{X}_t(r^*) = [\mathbf{Y}_t' \mathbf{I}(Y_{t-d} \leq r^*), \mathbf{Y}_t' \mathbf{I}(Y_{t-d} > r^*)]'$ , donde  $\mathbf{Y}_t = (1, Y_{t-1}, \dots, Y_{t-p})'$ . Entonces, el procedimiento para obtener los niveles de significación asintóticos del estadístico  $W_n$  será idéntico al detallado para el caso homocedástico, salvo que los números aleatorios  $\{u_t\}$  vendrán generados por una distribución  $N(0, e_t)$ .

### Contrastes de bondad de ajuste

Existen multitud de contrastes de bondad de ajuste en diferentes campos. Ling y Tong (2011) hacen una revisión de los más significativos en el campo de las series temporales. En esta sección se van a revisar dos propuestas que permiten contrastar lo adecuado que es el uso de un modelo SETAR.

**Contraste basado en contrastes del tipo Score** Ling y Tong (2011) proponen un contraste de bondad de ajuste general que permite comparar entre dos modelos paramétricos cualesquiera. Sea  $\hat{\theta}_n$  el estimador máximo verosímil de  $\theta_0$  bajo la hipótesis nula, se

supone que se cumple

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \Sigma^{-1} \sum_{t=1}^n \frac{D_t(\theta_0)}{\sqrt{n}} + o_p(1), \quad (2.53)$$

donde  $D_t(\theta_0)$  denota el *score* de  $\theta$  evaluado en  $\theta_0$  y  $\Sigma = E[D_t(\theta_0) D_t(\theta_0)']$  es la matriz de información.

El contraste propuesto por Ling y Tong (2011) se basa en el proceso empírico

$$T_n(r, \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n D_t(\theta_0) I(Y_{t-d} \leq r). \quad (2.54)$$

Entonces, estudiando el proceso (2.54) y observando el supuesto (2.53), se define la matriz de información  $\Sigma_r = E[D_t(\theta_0) D_t(\theta_0)' I(Y_{t-d} \leq r)]$ , y el valor  $A = \inf \{r : \Sigma = \Sigma_r\}$ , siendo generalmente  $A = \infty$ . Ling y Tong (2011) proponen entonces emplear el estadístico

$$S_n^a = \max_{a \leq r \leq A} \frac{[\beta' \hat{\Sigma}_{nr}^{-1} T_n(r, \hat{\theta}_n)]^2}{\beta' (\hat{\Sigma}_{na}^{-1} - \hat{\Sigma}_n^{-1}) \beta}, \quad (2.55)$$

siendo  $\beta$  un vector de constantes diferentes a cero de dimensión  $(p \times 1)$ . Además  $\hat{\Sigma}_{nr}$  y  $\hat{\Sigma}_n$  son los estimadores de  $\Sigma_r$  y de  $\Sigma$ , y se calculan mediante

$$\hat{\Sigma}_{nr} = \sum_{t=1}^n \frac{D_t(\hat{\theta}_n) D_t(\hat{\theta}_n)' I(Y_{t-d} \leq r)}{n}, \quad (2.56)$$

y

$$\hat{\Sigma}_n = \sum_{t=1}^n \frac{D_t(\hat{\theta}_n) D_t(\hat{\theta}_n)'}{n}. \quad (2.57)$$

Ling y Tong (2011) explican que la elección de  $\beta$  es un problema actualmente abierto, puesto que encontrar su óptimo llevaría a buscar la distribución asintótica de  $\max_{\beta} S_n^a$  en lugar de la de  $S_n^a$ , lo que complicaría el problema. En la práctica proponen emplear  $\beta = (1, \dots, 1)'$ . Como valor de  $a$  proponen emplear el cuantil 5p% de la muestra.

La aproximación a los valores críticos de  $S_n^a$  para rechazar la hipótesis nula con un nivel de significación  $\alpha$  se podrá calcular mediante  $P(S_n^a > C_\alpha)$ . Ling y Tong (2011) muestran que la constante  $C_\alpha$  es calculada empleando

$$P \left[ \max_{\tau \in [0,1]} B^2(\tau) \geq C_\alpha \right] = \alpha, \quad (2.58)$$



siendo  $B(\tau)$  un movimiento Browniano estándar. Entonces, dado que según Schorack y Wellner (1986, p.34),

$$P \left[ \max_{\tau \in [0,1]} B^2(\tau) \geq x \right] = 1 - \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp \left[ -\frac{(2k+1)^2 \pi^2}{8x} \right], \quad x > 0, \quad (2.59)$$

se puede obtener que  $C_{0.1} = 3.83$ ,  $C_{0.05} = 5.00$  y  $C_{0.01} = 7.63$ .

**Modelo paramétrico contra modelo no paramétrico** Generalmente los contrastes de bondad de ajuste comparan modelos no lineales con lineales, o con otros no lineales. Una opción interesante sería comparar el modelo no lineal con uno no paramétrico. Cai et al. (2000) muestran como un posible ejemplo el contraste de un modelo SETAR contra un modelo de coeficientes variables. Un modelo SETAR puede ser escrito como

$$Y_t = a_1(Y_{t-d}, \theta) Y_{t-1} + \dots + a_p(Y_{t-d}, \theta) Y_{t-p} + e_t, \quad (2.60)$$

donde  $\theta$  es un vector de parámetros desconocidos. Mientras que un modelo de coeficientes variables puede ser escrito en la forma general de un modelo autorregresivo funcional

$$Y_t = a_1(Y_{t-d}) Y_{t-1} + \dots + a_p(Y_{t-d}) Y_{t-p} + e_t. \quad (2.61)$$

Entonces, la suma de cuadrados residual bajo la hipótesis nula vendrá dada por

$$SCR_0 = \frac{1}{n} \sum_{t=p+1}^n \left\{ Y_t - a_1(Y_{t-d}, \hat{\theta}) Y_{t-1} - \dots - a_p(Y_{t-d}, \hat{\theta}) Y_{t-p} \right\}^2, \quad (2.62)$$

y bajo la hipótesis alternativa

$$SCR_1 = \frac{1}{n} \sum_{t=p+1}^n \left\{ Y_t - \hat{a}_1(Y_{t-d}) Y_{t-1} - \dots - \hat{a}_p(Y_{t-d}) Y_{t-p} \right\}^2. \quad (2.63)$$

De esta manera el estadístico vendrá definido por

$$T_n = \frac{SCR_0 - SCR_1}{SCR_1},$$

donde la hipótesis nula será rechazada para valores grandes de  $T_n$ . Cai et al. (2000) proponen un procedimiento Bootstrap para encontrar los valores críticos de  $T_n$ .

### Contrastes basados en el contraste de aditividad de Tukey

Tsay (1986) desarrolló un contraste general de no linealidad basado en el contraste de aditividad de Tukey (1949), mostrando además que era más potente que el propuesto por Keenan (1985). Por otra parte, Petrucci y Davies (1986) propusieron un contraste específico para modelos SETAR basándose en la idea de ordenar los datos respecto a la variable umbral, para así conseguir agrupar las observaciones pertenecientes a cada régimen de comportamiento. Finalmente, Tsay (1989) propuso un contraste para detectar modelos SETAR uniendo la idea de Petrucci y Davies (1986) con la de Tsay (1986).

El contraste propuesto por Tsay (1989) ha sido uno de los más empleados a lo largo de la literatura, debido, principalmente, a que el valor del umbral  $r$  no influye en el contraste, por lo que no es necesario asumir ningún tipo de estructura previa en el modelo SETAR contrastado. Esto se debe a que el modelo SETAR(k;p,d) de la hipótesis alternativa no depende del parámetro  $r$ , ya que Tsay (1989) escribe el modelo en forma de un proceso autorregresivo ordenado. Para ello la serie  $Y_1, \dots, Y_{n-d}$  es ordenada de menor a mayor obteniendo una nueva serie  $Y_{\pi_1}, \dots, Y_{\pi_{n-d}}$ , donde  $\pi_i$  contiene los índices del vector ordenado. Entonces, el modelo de la hipótesis alternativa en forma de autorregresivo ordenado sería

$$Y_{\pi_i+d} = \phi_0^{(k)} + \sum_{j=1}^p \phi_j^{(k)} Y_{\pi_i+d-j} + e_{\pi_i+d}^{(k)}, \quad (2.64)$$

siendo

$$k = \begin{cases} 1 & , \text{ si } i \leq s \\ 0 & , \text{ si } i > s \end{cases}, \quad (2.65)$$

donde, como puede verse no será necesario el conocimiento del umbral. La ordenación de una serie temporal puede no ser un paso muy intuitivo. El ordenamiento realizado asegura que los términos de la derecha en (2.64) son los retardos del término de la izquierda, por lo tanto las filas serán intercambiables si el proceso de las innovaciones es independiente de las observaciones. En la sección 3.1 se darán más detalles sobre el modelo autorregresivo ordenado y sus propiedades.

Empleando un proceso autorregresivo ordenado Tsay (1989) propuso el siguiente procedimiento para contrastar la existencia de un modelo SETAR.

1. Coger las primeras  $m$  observaciones de la serie ordenada ( $m = 30$  es lo recomendado por Tsay) y realizar por mínimos cuadrados la regresión siguiente,

$$Y_{\pi_i+d} = \beta_0 + \sum_{j=1}^p \beta_j Y_{\pi_i+d-j} + e_{\pi_i+d}, \quad i = 1, \dots, m. \quad (2.66)$$

De esta manera se estima el vector de coeficientes  $\hat{\beta}_m$ .

2. Calcular los residuos predictivos estandarizados un paso adelante mediante,

$$\hat{e}_{\pi_{m+1}+d} = Y_{\pi_{m+1}+d} - \hat{\beta}_{0,m} - \sum_{j=1}^p \hat{\beta}_{j,m} Y_{\pi_{m+1}+d-j}. \quad (2.67)$$

Incorporar el punto  $m + 1$  a la muestra y realizar el paso 1, actualizando el vector de coeficientes empleando mínimos cuadrados recursivos, obteniendo  $\hat{\beta}_{m+1}$ . Repetir de manera recursiva para todas las observaciones.

- (i) Realizar la regresión lineal

$$\hat{e}_{\pi_i+d} = \alpha_0 + \sum_{j=1}^p \alpha_j Y_{\pi_i+d-j} + v_t, \quad i = m + 1, \dots, n - d - m. \quad (2.68)$$

Entonces el estadístico del contraste se calcula mediante

$$\hat{F}_n = \left( \frac{X2 - X1}{X1} \right) \left( \frac{n - d - m - p}{p + 1} \right), \quad (2.69)$$

donde

$$\begin{aligned} X1 &= \sum_{t=m+1}^{n-d-m} \hat{v}_t^2, \\ X2 &= \sum_{t=1}^{n-p} \hat{e}_t^2, \end{aligned} \quad (2.70)$$

y la distribución límite de  $\hat{F}_n$  será una distribución F tal que

$$\hat{F}_n \sim F_{p+1, n-d-m-p}. \quad (2.71)$$

### 2.5.2 Algoritmos de búsqueda intensiva

El uso de contrastes nos dará una idea de si el uso del modelo SETAR es adecuado o no, pero no nos dará la forma de la partición del espacio de los estados en el modelo SETAR (2.2). La mayor parte de procedimientos existentes en la literatura que permiten la identificación de los umbrales, están basados en métodos de búsqueda intensiva.

Dada una colección  $\{\Omega_j, j = 1, \dots, J\}$  de posibles particiones del espacio de los estados, se calcula la función de costes  $S(\{\Omega_j\})$  que minimiza a (2.8) mediante la expresión (2.9), para todos los  $j = 1, \dots, J$ , es decir, se busca la partición  $\Omega_j$  que minimiza  $S(\{\Omega_j\})$ .

En la mayoría de los casos, el conjunto de posibles particiones  $\{\Omega_j\}$  será infinito, por lo que es necesario acotar las posibilidades. En la práctica se escoge un valor máximo de posibles regímenes  $k^{\max}$  (4 como mucho) y un rango dentro de la muestra de posibles valores de umbrales, por ejemplo, el 60 por ciento de la muestra.

### Procedimiento basado en criterios de selección

Al igual que la selección de los ordenes de un modelo, es bastante intuitivo pensar que el uso de criterios de selección puede ser útil para estimar los valores de los umbrales. El principal problema, es que en cuanto el tamaño muestral sea elevado el número de evaluaciones a realizar crecerá de manera exponencial. Wong y Li (1998) proponen un procedimiento basado en el propuesto por Tong (1990, p.379) para seleccionar el orden autorregresivo en cada régimen. El procedimiento es el siguiente:

- (i) Fijar los valores máximos del *delay*  $d$  y de los ordenes autorregresivos en cada régimen  $p_1^{\max}, \dots, p_k^{\max}$ .
- (ii) Asumir que el umbral pertenece a un intervalo de la muestra  $r \in I_r = [l, u] \subset \mathbb{R}$ . Los límites del intervalo suele ser los percentiles  $l = a\%$  y  $u = b\%$ , donde una elección adecuada suelen ser  $a = 25$  y  $b = 75$ , ó  $a = 10$  y  $b = 90$ . Hay que ser cuidadoso con la elección de un intervalo pequeño puesto que hay que garantizar que haya observaciones pertenecientes a todos los regímenes.
- (iii) Sean los candidatos a umbrales  $r_i$ ,  $i \in I_r$ , calcular el valor del criterio AICc y seleccionar aquellos valores del *delay*  $d$  y de los ordenes autorregresivos  $p_1, \dots, p_k$ , según

$$\min_{\substack{1 \leq p_j \leq p_j^{\max}; j=1, \dots, k \\ 1 \leq d \leq d^{\max}}} \text{AICc}([p_1, \dots, p_k]).$$

El valor mínimo encontrado se denota como  $\text{AICc}(r_i)$ , con sus correspondientes  $d(r_i), p_1(r_i), \dots, p_k(r_i)$ .

- (iv) Encontrar el valor mínimo de  $\text{AICc}(r_i)$ ,  $i \in I_r$  y seleccionar la estructura del modelo que minimice dicho criterio como el modelo SETAR adecuado.

### Procedimiento basado en validación cruzada

De Gooijer (2001) propone emplear los criterios de selección basados en la idea de validación cruzada que mostramos anteriormente para desarrollar un procedimiento que permite modelizar modelos SETAR. El procedimiento propuesto es el siguiente:

- (i) Fijar el máximo valor del parámetro *delay*  $d_{\max}$ , el número de regímenes  $k$  y los ordenes autorregresivos máximos  $p_1^{\max}, \dots, p_k^{\max}$  que podrán ser seleccionados en cada régimen del modelo SETAR.
- (ii) Definir  $R$  candidatos a ser los valores de los umbrales. Estos valores deben estar dentro del intervalo  $[l, u]$ , donde  $l$  y  $u$  son cuantiles del proceso  $\{Y_t\}$ .
- (iii) Seleccionar  $K$  posibles valores de los ordenes autorregresivos para cada régimen, teniendo como orden máximo en cada régimen  $p_1^{\max}, \dots, p_k^{\max}$ . De esta manera para un valor del *delay*  $d$  existen  $G$  modelos SETAR a probar, siendo  $G = K^k R$ . Es decir, si se escogen como valores de los umbrales el 50% de una muestra con 200 observaciones y se fijan como ordenes autorregresivos máximos  $p = 3$ , el número de modelos a evaluar sería 900 para cada valor de  $d$ . En cuanto el modelo que se quiere evaluar sea un poco complejo el número de evaluaciones crecerá de manera exponencial.
- (iv) Estimar los parámetros del  $g$ -ésimo modelo SETAR omitiendo  $m$  observaciones y empleando el método de mínimos cuadrados sobre el modelo autorregresivo ordenado. Así se obtiene  $\hat{\phi}_i$  que es la estimación de los diferentes parámetros autorregresivos de los diferentes regímenes empleando la  $i$ -ésima muestra de calibración.
- (v) Sea  $\{\hat{Y}_{t,i}^g\}$  la predicción un paso adelante del proceso  $Y_t$  empleando el  $g$ -ésimo modelo y la  $i$ -ésima muestra de validación, calcular los residuos predictivos

$$\hat{e}_t \left( \hat{\theta}_i (\text{Modelo}_g) \right) = Y_t - \hat{Y}_{t,i}^g, t = 1, \dots, n; i = 1, \dots, L; g = 1, \dots, G,$$

con  $L = \lceil n/m \rceil$ . Entonces, para cada uno de los modelos que hay que evaluar se tienen que realizar  $L$  estimaciones. Dado que el valor de  $m$  que recomiendan los autores es  $m = 1$ , entonces, volviendo al ejemplo que mostramos en el paso (iii),  $L = 200$ , y el número de estimaciones que hay que realizar para cada valor de  $d$  que se desea probar asciende a 1800.

- (vi) Calcular cualquiera de los criterios de selección mostrados en la sección 2.4.3 y elegir como modelo SETAR aquella estructura  $M_g$  que minimice el valor del criterio. De Gooijer (2001) aconseja emplear el criterio de selección basado en validación cruzada insesgado (2.23) según los resultados obtenidos en las simulaciones.

## 2.6 Predicción con modelos SETAR

El cálculo de predicciones empleando modelos SETAR tiene una ventaja respecto a otro tipo de modelos no lineales, y es que al ser un modelo lineal a trozos realizar predicciones dentro de cada régimen es sencillo. Para ello es necesario conocer a que régimen pertenece la variable en el instante de tiempo a predecir, es decir, necesitamos conocer el valor de la variable umbral  $Y_{t-d}$ . El problema surge cuando el horizonte de predicción  $h$  es mayor que el parámetro *delay*  $d$ , puesto que en ese caso, la variable umbral deja de ser observada, por lo que calcular las predicciones de manera exacta es complicado. Existe la creencia de que calcular predicciones exactas en este caso es imposible, cosa que Tong (1990, p. 347) desmiente. No es imposible, pero si es muy complejo.

Un modelo SETAR puede ser escrito en forma de un modelo autorregresivo no lineal

$$Y_t = h(Y_{t-1}) + e_t, \quad (2.72)$$

donde  $\{e_t\}$  es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza finita. Entonces, calcular la predicción a horizonte  $h$  significa estimar la esperanza condicionada  $E[Y_{t+h}|Y_t]$ . Calcular dicha esperanza condicionada directamente es complejo, por ello Tong (1990, p.346) propone una estimación empleando la distribución predictiva, es decir, la distribución condicionada de la predicción. Dicha densidad predictiva puede ser estimada empleando la relación de Chapman-Kolmogorov

$$f(Y_{t+h}|Y_t) = \int_{-\infty}^{\infty} f(Y_{t+h}|Y_{t+1}) f(Y_{t+1}|Y_t) dY_{t+1}. \quad (2.73)$$

De esta manera, empleando métodos de integración numérica (ver, p.ej., Tong, 1990 Cap. 4.2.4) se puede estimar la media y la varianza condicionada. El problema principal es que este tipo de métodos numéricos puede ser complejo y computacionalmente costoso dependiendo de la complejidad de la estructura del modelo SETAR utilizado.

Granger y Teräsvirta (1993) proponen cinco técnicas para calcular predicciones a horizonte  $h$  empleando modelos autorregresivos no lineales sin necesidad de emplear técnicas de integración numérica. Por simplicidad en la notación durante esta sección asumiremos un modelo SETAR(2;1,1), además la predicción será  $\hat{Y}_{t+h|t} = E[Y_{t+h}|Y_t]$ . Entonces los métodos propuestos son:

- Método de errores de predicción normales propuesto por Al-Qassam y Lane (1989) para su uso en modelos EAR y adaptado a modelos SETAR por De Gooijer y De Bruin (1998). La predicción se estima mediante una ponderación de las predicciones correspondientes a cada uno de los regímenes del modelo SETAR em-

pleado. De Gooijer y De Bruin (1998) muestran que para un modelo SETAR(2;1,1) la predicción se calcula mediante

$$\begin{aligned} \hat{Y}_{t+h|t} = & \Pi \left( \frac{\hat{r} - \hat{Y}_{t+h-1|t}}{\hat{\sigma}_{n+h-1}} \right) \hat{Y}_{t+h|t}^{(1)} + \Pi \left( -\frac{\hat{r} - \hat{Y}_{t+h-1|t}}{\hat{\sigma}_{n+h-1}} \right) \hat{Y}_{t+h|t}^{(2)} \\ & + \left( \hat{\phi}_1^{(2)} - \hat{\phi}_1^{(1)} \right) \hat{\sigma}_{n+h-1} \pi \left( \frac{\hat{r} - \hat{Y}_{t+h-1|t}}{\hat{\sigma}_{n+h-1}} \right), \end{aligned} \quad (2.74)$$

donde  $(\hat{\phi}_1^{(1)}, \hat{\phi}_1^{(2)})$  y  $(\hat{Y}_{t+h|t}^{(1)}, \hat{Y}_{t+h|t}^{(2)})$  son la estimación de los parámetros y la predicción para cada régimen.  $\pi(\cdot)$  y  $\Pi(\cdot)$  son, respectivamente, las funciones de densidad y de distribución de una Normal con media 0 y varianza 1. Por ello, este método solo será útil en el caso de que los residuos sigan una distribución Normal. El valor  $\hat{\sigma}_i$  es la desviación del error de predicción  $i$ -ésimo.

- Método de predicción adaptativa. La idea es emplear diferentes estimaciones de los parámetros para diferentes horizontes de predicción, como sugirió Cox (1961). En series temporales esta aproximación ha sido empleada para su uso en modelos lineales por Findley (1983, 1985) y Tiao y Tsay (1994). Para calcular la predicción a horizonte  $h$ , el modelo es estimado minimizando la suma de los cuadrados de los errores de predicción dentro de la muestra a horizonte  $h$ . Para ilustrar más claramente vamos a suponer un modelo AR(1). En ese caso, la predicción a horizonte  $h$  se calculará

$$\hat{Y}_{t+h|t} = \phi_0 + \phi_1 \hat{Y}_{t+h-1|t}. \quad (2.75)$$

Entonces, empleando el método de predicción adaptativa los parámetros del modelo (2.75) serán estimados minimizando la suma de cuadrados de los errores de predicción a horizonte  $h$ , es decir, minimizando

$$\sum_{t=2}^{n-h} e_{t+h|t}^2 = \sum_{t=2}^{n-h} \left[ Y_{t+h} - \hat{Y}_{t+h|t} \right]^2, \quad (2.76)$$

lo que nos lleva a un método de estimación por mínimos cuadrados no lineales. Clements y Smith (1997) muestran como calcular las predicciones para un modelo SETAR, aunque lo llaman método de estimación dinámica.

- Métodos de Monte Carlo. Dado un modelo SETAR, es claro que la predicción a horizonte  $h \leq d$  puede ser calculada sin ningún tipo de complicación, puesto que la variable umbral siempre será observada. Cuando el horizonte sea  $h > d$  las predicciones serán calculadas simulando diferentes trayectorias posibles, que luego serán empleadas para estimar la predicción. Entonces, sea  $\hat{Y}_{t+d|t}$  la última

predicción que puede ser calculada de manera exacta, la  $j$ -ésima trayectoria a partir del instante de tiempo  $d$  será generada empleando

$$\hat{Y}_{t+d+1|t}^j = \left( \phi_0^{(i)} + \phi_1^{(i)} \hat{Y}_{t+d|t}^j + \xi^{(i)} \right) \mathbf{I} \left( \hat{Y}_{t+d|t}^j \in \Omega_i \right), \quad i = 1, 2; \quad (2.77)$$

y

$$\hat{Y}_{t+k|t}^j = \left( \phi_0^{(i)} + \phi_1^{(i)} \hat{Y}_{t+d+1|t}^j + \xi^{(i)} \right) \mathbf{I} \left( \hat{Y}_{t+d+1|t}^j \in \Omega_i \right), \quad i = 1, 2; \quad k = d+2, \dots, h, \quad (2.78)$$

siendo  $\xi^{(i)}$  un número aleatorio generado a partir de la densidad correspondiente al  $i$ -ésimo régimen. Una vez computadas las  $J$  trayectorias la predicción puede ser calculada mediante la media de dichas trayectorias

$$\hat{Y}_{t+k|t} = \frac{1}{J} \sum_{j=1}^J \hat{Y}_{t+k|t}^j, \quad k = d+1, \dots, h. \quad (2.79)$$

Este método puede ser implementado fácilmente para cualquier tipo de modelos y evita tener que asumir que la distribución es Normal.

- Métodos Bootstrap. El funcionamiento es igual al detallado para el método de Monte Carlo, salvo en la generación del número aleatorio  $\xi^{(i)}$ . En este caso, en vez de emplear una distribución, se realiza un muestreo con reemplazamiento entre los residuos del  $i$ -ésimo régimen obtenidos en la estimación del modelo.
- Método Skeleton. En este caso no se realiza ningún tipo de simulación, sino que las predicciones se van calculando mediante

$$\hat{Y}_{t+k|t}^j = \left( \phi_0^{(i)} + \phi_1^{(i)} \hat{Y}_{t+d+1|t}^j \right) \mathbf{I} \left( \hat{Y}_{t+d+1|t}^j \in \Omega_i \right), \quad i = 1, 2; \quad k = d+2, \dots, h, \quad (2.80)$$

es decir, asumiendo que los valores que se van calculando se convierten en la variable umbral y descartando la existencia del error debido al uso como variable umbral de una estimación y no de un valor observado.

En Clements y Smith (1997) podemos ver un estudio comparativo de las diferentes técnicas para modelos SETAR. Los autores concluyen que, en general, el método más eficiente y flexible es el método de Monte Carlo.



## Capítulo 3

# Identificación de modelos SETAR mediante estimación recursiva

En este capítulo se describe el procedimiento de identificación de modelos SETAR propuesto, y se comprueba su eficacia en muestras finitas mediante simulaciones. La herramienta propuesta ha recibido el nombre de ARLS (*Arranged Recursive Least Squares*). En dicho trabajo se propone un método de identificación basandonos en dos ideas básicas, la ordenación del proceso respecto a la variable que provoca la no linealidad como ya hicieron Petrucci y Davies (1986) y Tsay (1989) y el uso de la estimación recursiva para así poder detectar un posible cambio estructural en la secuencia de estimaciones. El contenido de este capítulo ha sido recientemente aceptado para su publicación en la revista *Journal of Forecasting* (Bermejo et al., 2011).

### 3.1 Modelo autorregresivo ordenado con parámetros variables en el tiempo

Sea  $Y_t$  un proceso autorregresivo con parámetros variables en el tiempo (TV-AR, *Time-Varying Autoregressive*) de orden  $p$  vendrá definido mediante

$$Y_t = X_t' \phi_t + e_t; \quad t = h, \dots, n; \quad (3.1)$$

donde el proceso  $\{e_t\}$  es una secuencia de variables aleatorias independientes e idénticamente distribuidas, tales que  $E[e_t] = 0$  y  $E[e_t^2] = \sigma_t^2 < \infty$ . El vector  $X_t = (1, Y_{t-1}, \dots, Y_{t-p})'$  es una muestra de variables explicativas.

Se define  $S$  como el conjunto de todas las posibles ordenaciones del vector temporal  $t = 1, \dots, n$ . Entonces  $s_{i,t}$  es la  $t$ -ésima posición, con  $t = 1, \dots, n$ , del  $i$ -ésimo

elemento de  $S$ , con  $i = 1, \dots, n!$ . El subíndice  $i$  será en ocasiones omitido cuando nos estemos refiriendo a un elemento genérico de  $S$ , por motivos de sencillez en la notación. Denotaremos con  $\pi_t$  a la  $t$ -ésima posición para el caso particular correspondiente a ordenar las observaciones en orden ascendente respecto a la variable umbral  $Y_{t-d}$ . Es decir,  $\pi_t$  es el instante de tiempo del  $t$ -ésimo menor elemento de  $(Y_h, \dots, Y_{n-d})$ , donde  $h = \max(1, p + 1 - d)$ . Entonces, la serie temporal ordenada  $Y_{s_t}$  seguirá un proceso autorregresivo ordenado de orden  $p$  con parámetros variables (TV-AAR( $p$ ), *Time-Varying Arranged Autoregression*) si tiene la forma

$$Y_{s_t} = X'_{s_t} \phi_{s_t} + e_{s_t}; \quad t = h, \dots, n; \quad (3.2)$$

donde, para cualquier ordenación  $s_t$ , perteneciente al conjunto de ordenaciones  $S$ ,  $e_{s_t}$  es una secuencia de variables aleatorias independientes e idénticamente distribuidas tales que  $E[e_{s_t}] = 0$  and  $E[e_{s_t}^2] = \sigma_t^2 < \infty$ .

Para ilustrar el modelo autorregresivo ordenado mostramos un ejemplo sencillo. Sea  $\{Y_t\}$  un proceso TV-AR(1)  $Y_t = \phi_{1t} Y_{t-1} + e_t$ , si ordenamos los pares  $(Y_t, Y_{t-1})$  usando la variable umbral  $Y_{t-d}$ , obtendremos el modelo TV-AAR(1)

$$\begin{pmatrix} Y_{\pi_h} \\ Y_{\pi_{h+1}} \\ \vdots \\ Y_{\pi_n} \end{pmatrix} = \begin{pmatrix} \phi_{1\pi_h} Y_{\pi_h-1} \\ \phi_{1\pi_{h+1}} Y_{\pi_{h+1}-1} \\ \vdots \\ \phi_{1\pi_n} Y_{\pi_n-1} \end{pmatrix} + \begin{pmatrix} e_{\pi_h} \\ e_{\pi_{h+1}} \\ \vdots \\ e_{\pi_n} \end{pmatrix}. \quad (3.3)$$

Es importante mostrar que la ordenación, bajo unos supuestos, no va a afectar a la estructura temporal de la variable. La idea de realizar ordenaciones en las variables ha sido empleada en la realización de contrastes de permutación, que fueron introducidos por Fisher (1935) y Pitman (1938). En este caso el único requisito que se le pide a la variable aleatoria para poder ser ordenada es que cumpla el principio de intercambiabilidad, introducido por De Finetti (1930).

**Definición 3.1 (Intercambiabilidad).** Sea  $s_t$  un vector de índices que posee una ordenación aleatoria cualquiera. Sea  $S$  el conjunto de todas las ordenaciones posibles que se pueden realizar sobre la variable aleatoria. Diremos que un vector aleatorio  $\mathbf{Y}_t$  es intercambiable si y solo si cualquier vector aleatorio  $\mathbf{Y}_{s_t}$  tiene la misma distribución multivariante, para todo  $s_t \in \mathbb{R}$ .

La intercambiabilidad requiere un supuesto muy rígido que será muy difícil de comprobar cuando se trabaje con series temporales. Por ello, en la práctica es suficiente con la intercambiabilidad de segundo orden.

**Definición 3.2 (Intercambiabilidad de segundo orden).** Un vector aleatorio  $\mathbf{Y}_t$  será intercambiable de segundo orden si para cualquier vector aleatorio  $\mathbf{Y}_{s_t}$  se cumple

$$\begin{aligned} E[\mathbf{Y}_{s_t}] &= \mu; \\ \text{Var}[\mathbf{Y}_{s_t}] &= \Sigma, \end{aligned}$$

para todo  $s_t \in \mathbb{R}$ .

Entonces, como mostró Wedlin (1997) se puede decir que la variable aleatoria  $Y_t$  será intercambiable de segundo orden siempre que sea estacionaria en sentido débil. Es decir, los modelos autorregresivos ordenados mantienen la propiedad de intercambiabilidad de segundo orden, en el sentido de que el proceso de los términos de error  $e_t$  de cualquier ordenación realizada mantendrá su matriz de covarianza sin cambios.

## 3.2 Estimación recursiva de de parámetros variables en el tiempo

Sea  $Y_{s_t}$  un proceso TV-AAR( $p$ ) definido en (3.1), será necesario emplear un método de estimación recursiva para calcular los estimadores de los parámetros variables en el tiempo. Existen varios algoritmos que permiten hacerlo, como son mínimos cuadrados recursivos (RLS, *Recursive Least Squares*) o el filtro de Kalman. El inconveniente del filtro de Kalman es que para emplearlo se necesita conocer el modelo que sigue la variación del parámetro. Además una de las ventajas del método RLS es que es especialmente útil cuando los cambios en los parámetros ocurren lentamente a lo largo del índice en el que estén ordenadas las observaciones. Es decir, los cambios se pueden producir de manera temporal, o como será nuestro caso, se producirán debido a que la variable que puede producir dichos cambios en el parámetro está ordenada.

### 3.2.1 Mínimos cuadrados recursivos y ponderados

El estimador de mínimos cuadrados ponderados (WLS, *Weighted Least Squares*)  $\hat{\phi}_t$  del modelo (3.1) es aquel que minimiza la función de costes

$$\hat{\phi}_t = \arg \min_{\phi} C_t(\phi), \quad (3.4)$$

donde

$$C_t(\phi) = \sum_{j=1}^t \kappa(t, j) (Y_t - X_t' \phi)^2, \quad (3.5)$$

siendo  $\kappa(t, t) = 1 \geq \kappa(t, t-1) \geq \dots \geq \kappa(t, 1) \geq 0$  una secuencia de pesos conocida como perfil de olvido. Cada una de estas cantidades  $\kappa(t, j)$  denotan el peso dado al  $j$ -ésimo residuo en la función de costes en el instante de tiempo  $t$ . A lo largo de este trabajo se emplea un perfil de olvido del tipo

$$\kappa(t, j) = \prod_{i=j+1}^t \lambda_i, \quad j < t, \quad (3.6)$$

donde  $\kappa(t, t) = 1$  and  $0 \leq \lambda_t \leq 1$  es el factor de olvido, que puede ser constante,  $\lambda_t = \lambda$ , o variable con el tiempo. El factor de olvido causa una reducción progresiva en la importancia de los datos pasados en la estimación. Por esta razón, la estimación es adaptativa con el tiempo.

El estimador WLS del modelo (3.1) es

$$\hat{\phi}_t = (\mathbf{X}_t' \mathbf{\Lambda}_t \mathbf{X}_t)^{-1} \mathbf{X}_t' \mathbf{\Lambda}_t \mathbf{Y}_t, \quad (3.7)$$

donde  $\mathbf{X}_t$  es la matriz  $(X_h, X_{h+1}, \dots, X_t)'$ ,  $\mathbf{\Lambda}_t$  es una matriz diagonal con los elementos  $\lambda_h, \lambda_h \lambda_{h+1}, \dots, \lambda_h \lambda_{h+1} \dots \lambda_t$  en la diagonal principal y el vector  $\mathbf{Y}_t = (Y_{s_h}, \dots, Y_{s_t})'$ . El estimador (3.7) puede ser calculado de manera recursiva empleando (ver, p.ej., Ljung and Söderström 1983)

$$\hat{\phi}_t = \hat{\phi}_{t-1} + M_t^{-1} X_t \hat{e}_t, \quad (3.8)$$

donde  $\hat{e}_t = Y_t - X_t' \hat{\phi}_{t-1}$  es el error de predicción un paso adelante y  $M_t = (X_t' \mathbf{\Lambda}_t X_t)$  es la matriz de ganancia que mide la dispersión los parámetros estimados. La expresión (3.8) es el estimador RLS. La matriz de ganancia  $M_t$  se calcula de manera recursiva mediante

$$M_t = \lambda_t M_{t-1} + X_t X_t'. \quad (3.9)$$

El problema es que realizar el cálculo de una matriz inversa en cada iteración ralentizará el algoritmo e incrementará los problemas numéricos. Para evitar este cálculo se halla la forma recursiva de calcular  $M_t$ , para lo que será necesario emplear el lema de inversión de matrices (ver, p.ej., Bodewig 1956).

**Lema 3.3 (Lema de Inversión de Matrices).** Sean  $B$  y  $D$  dos matrices no singulares y de orden  $m$  y  $n$ , respectivamente y sea  $C$  una matriz de orden  $n \times m$ . Entonces,

$$(B + C')^{-1} = B^{-1} - B^{-1} C' [C B^{-1} C']^{-1} C B^{-1}.$$

Empleando el lema 3.3, se calcula la inversa de la matriz de ganancias de manera

recursiva mediante

$$M_t^{-1} = \frac{1}{\lambda_t} \left( M_{t-1}^{-1} - \frac{M_{t-1}^{-1} X_t X_t' M_{t-1}^{-1}}{\lambda_t + X_t' M_{t-1}^{-1} X_t} \right), \quad (3.10)$$

donde sólo será necesario invertir  $\lambda_t + X_t' M_{t-1}^{-1} X_t$  que es un escalar.

### 3.2.2 Factores de olvido adaptativos

La correcta elección del factor de olvido será clave para que el método funcione eficientemente, puesto que de dicha elección dependerá obtener una buena estimación de los parámetros del modelo y, por lo tanto, la eficiencia de las predicciones. Como dijimos anteriormente el factor de olvido puede ser constante o variable. Puede demostrarse (Haykin et al., 1997) que emplear RLS con un factor de olvido  $\lambda_t$  constante es muy similar a usar el filtro de Kalman asumiendo un paseo aleatorio para  $\phi_t$ . Por lo tanto, si no se quiere imponer dicha evolución, se debe usar un  $\lambda_t$  variable, para que la evolución venga dictada por los datos.

Para comprobar mejor la influencia del factor de olvido  $\lambda_t$  en la estimación se desarrolla la expresión (3.5)

$$\begin{aligned} C_t(\phi) &= (y_t - X_t' \phi)^2 + \lambda_t C_{t-1}(\phi) \\ &= \dots \\ &= (y_t - X_t' \phi)^2 + \lambda_t (y_{t-1} - X_{t-1}' \phi)^2 + \lambda_t \lambda_{t-1} (y_{t-2} - X_{t-2}' \phi)^2 \\ &\quad + \dots + \lambda_t \lambda_{t-1} \dots \lambda_2 (y_1 - X_1' \phi)^2. \end{aligned} \quad (3.11)$$

Es fácil ver que al tomar  $\lambda_t$  valores entre 0 y 1, el efecto de las observaciones antiguas irá decreciendo de manera exponencial. De esta manera un valor alejado de 1 da una mayor dependencia de los datos recientes y permite notar más rápidamente cambios en la estimación. El problema es que esto también provoca una mayor variabilidad en la estimación, pues como puede verse en la expresión (3.10), que como dijimos es una medida de la dispersión de la estimación, es claro que el valor de  $M_t^{-1}$  crecerá cuando  $\lambda_t$  decrezca. La dificultad, por lo tanto, estribará en encontrar un valor compensado.

Algunos de los factores de olvido variables que se han propuesto son:

- Factor de olvido basado en los errores de predicción (Fortescue et al., 1981). Proponen emplear como factor de olvido un ratio entre el error de predicción observado respecto al esperado. Se calcula según

$$\lambda_t^{\text{pre}} = 1 - \alpha \frac{\hat{e}_t^2}{1 + X_{s_t}' M_{s_{t-1}}^{-1} X_{s_t}}, \quad (3.12)$$

donde  $\alpha$  es un parámetro definido por el usuario. Dicho parámetro es una de las dificultades para el uso de este método, pues no existe modo formal de selección. Observando la expresión (3.12) puede verse la existencia de la expresión  $X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t}$ . Es sabido, según la teoría clásica de regresión, que  $X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t}$  es una medida de la influencia de la  $t$ -ésima observación en la muestra de datos, o lo que es lo mismo, es una medida de la distancia de  $X_t$  al centro de gravedad de los datos. Entonces, si las nuevas observaciones influyen poco en la muestra, pero tienen un error de predicción grande, se obtendrá un valor de  $\lambda_t^{\text{pre}}$  pequeño y detectará el cambio en los parámetros que se produce. Sin embargo, si además de un error de predicción grande, las nuevas observaciones provocan cambios en el centro de gravedad de los datos, el  $\lambda_t^{\text{pre}}$  obtenido será insensible a los cambios en los parámetros (ver Sánchez, 2006a).

- Factor de olvido basado en la influencia de las nuevas variables (Landau et al., 1998). El factor de olvido está definido por

$$\lambda_t^{\text{lev}} = 1 - \frac{X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t}}{1 + X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t}}, \quad (3.13)$$

es decir, el ratio se basa en la distancia de las nuevas observaciones sobre las restantes. Para estudiar el comportamiento de  $\lambda_t^{\text{lev}}$ , se premultiplicar y postmultiplicar la expresión (3.10) por  $X_{s_t}$ , obteniendo

$$\begin{aligned} X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t} &= \frac{1}{\lambda_t} \left( X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t} - \frac{X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t} X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t}}{\lambda_t + X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t}} \right) \\ &= \frac{X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t}}{\lambda_t + X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t}} \equiv h_t(\lambda_t) \end{aligned} \quad (3.14)$$

Así, si se descarta toda la información previa, entonces  $\lambda_t = 0$  y por lo tanto,  $h_t(0) = 1$ , siendo la influencia de las nuevas observaciones máxima. Por el contrario, si no se aplica ningún olvido en la estimación,  $\lambda_t = 1$ , y

$$h_t(1) = \frac{X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t}}{1 + X'_{s_t} M_{s_{t-1}}^{-1} X_{s_t}}, \quad (3.15)$$

será la mínima influencia que se puede obtener de los datos. Entonces la expresión (3.13) puede ser escrita como

$$\lambda_t^{\text{lev}} = 1 - h_t(1). \quad (3.16)$$

El factor de olvido de Landau et al. (1998) será adecuado cuando un cambio en

los parámetros va asociado a un cambio en el centro de gravedad provocado por las nuevas observaciones, pero será insensible en caso de que dicho cambio en el centro de gravedad no exista.

- Factor de olvido basado en la distancia de Cook (Sánchez, 2006a). El factor de olvido se calcula mediante

$$\lambda_t^{\text{Cook}} = \lambda_{\min} + (1 - \lambda_{\min}) P(\chi_m^2 > mD_t), \quad (3.17)$$

donde  $\lambda_{\min}$  es la cota inferior del factor de olvido especificada por el usuario,  $m$  es el número de parámetros en (3.1) y  $D_t$  es una versión variable con el tiempo de la distancia de Cook calculada según

$$D_t = \frac{X'_{s_t} M_{s_t-1}^{-1} X_{s_t} \hat{e}_{s_t}^2}{m \hat{\sigma}_{s_t-1}^2 (1 + X'_{s_t} M_{s_t-1}^{-1} X_{s_t})}, \quad (3.18)$$

donde  $\hat{\sigma}_{s_t-1}^2$  es una estimación de  $\sigma^2$ . Sánchez (2006a) muestra que este factor de olvido combina las ventajas de (3.12) y (3.13), lo que es una ventaja pues, como hemos visto, ambos métodos funcionan bien donde falla el otro. La distancia de Cook es una medida de la influencia de las observaciones sobre los parámetros estimados, por lo que, si se obtiene una distancia pequeña, significará que las nuevas observaciones influyen poco y que, por lo tanto, el factor de olvido debería de ser mayor para tener en cuenta a los datos pasados. En caso contrario, si se obtiene una distancia alta, significará una gran influencia de las nuevas observaciones y un valor pequeño del factor de olvido para así descartar la información previa.

### 3.2.3 Propiedades del estimador RLS con factor de olvido

Las propiedades del estimador RLS con un factor de olvido variable son complejas. La distribución del estimador para un modelo de regresión variable con el tiempo son desconocidas. Sin embargo, para nuestro procedimiento de identificación de modelos SETAR se necesitan las propiedades del estimador RLS bajo el supuesto de un proceso autorregresivo invariante en el tiempo.

Para un proceso autorregresivo sin olvido, es decir, con  $\lambda_t = 1$ , el error cuadrático medio (ECM) del estimador de mínimos cuadrados ordinarios es (Fuller y Hasza, 1985; Kunitomo y Yamamoto, 1985)

$$\text{ECM}[\hat{\phi}_{\text{OLS}}] = E \left[ \left( \hat{\phi} - \phi \right) \left( \hat{\phi} - \phi \right)' \right] = \frac{\sigma^2}{n} \Gamma^{-1} + O(n^{-3/2}), \quad (3.19)$$

donde  $\Gamma = E(X_t X_t')$ , que puede ser estimado mediante  $\hat{\Gamma} = n^{-1} (X_t' X_t)^{-1}$ . El uso de un factor de olvido puede ser interpretado como una reducción del tamaño muestral, ya que empleando mínimos cuadrados ordinarios cada observación tiene el mismo peso en la estimación, sin embargo en el estimador (3.8), el tamaño muestral efectivo es menor que  $n$ . Si usamos, por simplicidad, un factor de olvido constante, el tamaño muestral efectivo es  $n_{eq} = 1 + \lambda + \dots + \lambda^{n-1}$ . Si  $n \rightarrow \infty$  el tamaño muestral efectivo es habitualmente conocido como memoria asintótica y se puede calcular fácilmente mediante

$$N_0 = \frac{1}{1 - \lambda}. \quad (3.20)$$

Consecuentemente, el error cuadrático medio del estimador RLS es mayor cuando  $\lambda$  es menor, puesto que es como si redujeramos el tamaño muestral. El error cuadrático medio asintótico para el estimador RLS con olvido puede ser aproximado mediante

$$\text{ECM}[\hat{\phi}_{\text{RLS}}] = \sigma^2 E[(X_t' \Lambda_t X_t)^{-1}] + O\left[(1 - \lambda)^{-3/2}\right], \quad (3.21)$$

por lo que si  $\lambda$  está próximo a 1 podrá ser aproximado como

$$\text{ECM}[\hat{\phi}_{\text{RLS}}] = \sigma^2 E[(X_t' \Lambda_t X_t)^{-1}], \quad (3.22)$$

y podrá ser estimado mediante

$$\widehat{\text{ECM}}[\hat{\phi}_{\text{RLS}}] = \hat{\sigma}_t^2 (X_t' \Lambda_t X_t)^{-1}, \quad (3.23)$$

donde  $\hat{\sigma}_t^2$  es un estimador de  $\sigma^2$ , por ejemplo, el estimador recursivo definido por

$$\hat{\sigma}_t^2 = \hat{\sigma}_{t-1}^2 + \frac{1}{t - p} (\hat{e}_t^2 - \hat{\sigma}_{t-1}^2). \quad (3.24)$$

### 3.3 Herramienta de identificación, ARLS

En esta sección se describe el procedimiento de identificación de modelos SETAR que propuesto, comprobando su eficacia en muestras finitas mediante simulaciones. La herramienta que propuesta recibe el nombre de ARLS (*Arranged Recursive Least Squares*).

La idea principal del método es la estimación del modelo TV-AAR(p) descrito en (3.2) usando un método de estimación recursiva, como el descrito en (3.8). Entonces, si el modelo autorregresivo es ordenado respecto a la variable umbral  $Y_{t-d}$ , se espera que la secuencia de estimaciones  $\hat{\phi}_{\pi_t}$  tienda a mostrar un cambio estructural en torno a el valor del umbral  $r$ . El procedimiento propuesto incluye una comprobación de que dicho cambio estructural es significativo.



A lo largo de esta sección, y por simplicidad en la notación, se asume un modelo SETAR(2;1, $d$ ) definido por

$$Y_t = (\phi + \delta \mathbf{I}_{(Y_{t-d} > r)}) Y_{t-1} + e_t. \quad (3.25)$$

### 3.3.1 Estimación inicial de la recursividad

Asumimos que  $d$  es conocido, ya que en la práctica se pueden probar diferentes valores y luego seleccionar el más adecuado empleando un criterio de selección de modelos. Así pues, una vez ordenados los datos, para realizar la estimación recursiva se necesita una estimación inicial  $\hat{\phi}_0$  para inicializar (3.8). Dicha estimación inicial va a ser un paso crucial a la hora del correcto funcionamiento del procedimiento, ya que este nos ayudará en la identificación del umbral, puesto que si  $\hat{\phi}_0$  es demasiado elevado o bajo el cambio estructural puede no ser detectado a tiempo.

La estimación inicial propuesta será calculada empleando mínimos cuadrados ordinarios con la muestra completa. De esta manera, la inicialización de la secuencia de estimaciones será un valor intermedio entre ambos regímenes, así la estimación recursiva  $\hat{\phi}_{\pi_t}$  tenderá a tomar la forma de un codo alrededor del valor del umbral  $r$ , y podrá ser identificado. Vamos a comprobar que con una estimación inicial  $\hat{\phi}_0$  calculada mediante OLS conseguimos siempre un valor intermedio.

**Proposición 3.4 (Estimación inicial).** Definimos  $\{Y_t\}$  como una serie temporal compuesta por  $Y_t^{(1)}$  y  $Y_t^{(2)}$ , es decir,  $Y_t = [Y_t^{(1)}, Y_t^{(2)}]'$ , donde  $\{Y_t^{(1)}\}$  y  $\{Y_t^{(2)}\}$  son dos procesos AR(1) con parámetros  $\phi$  y  $\phi + \delta$ , respectivamente. Análogamente definimos  $X_t = [Y_{t-1}^{(1)}, Y_{t-1}^{(2)}]'$ . Entonces, sean  $M_t^{(1)} = (Y_{t-1}^{(1)} Y_{t-1}^{(1)})^{-1}$  y  $M_t^{(2)} = (Y_{t-1}^{(2)} Y_{t-1}^{(2)})^{-1}$ , si ajustamos un modelo AR(1) a  $Y_t$  usando mínimos cuadrados ordinarios, obtendremos

$$\hat{\phi}_0 = \hat{\omega} \hat{\phi} + (1 - \hat{\omega})(\widehat{\phi + \delta}),$$

siendo  $\hat{\omega} \in (0, 1)$ .

*Demostración.* La estimación por mínimos cuadrados ordinarios de un proceso AR(1) será  $\hat{\phi}_0 = (X_t' X_t)^{-1} X_t' Y_t$ , o lo que es lo mismo empleando la composición de la serie temporal  $Y_t$

$$\hat{\phi}_0 = \left( Y_{t-1}'^{(1)} Y_{t-1}^{(1)} + Y_{t-1}'^{(2)} Y_{t-1}^{(2)} \right)^{-1} \left( Y_{t-1}'^{(1)} Y_t^{(1)} + Y_{t-1}'^{(2)} Y_t^{(2)} \right).$$

Ahora dado que  $(A + B)^{-1} = B^{-1} (A^{-1} + B^{-1})^{-1} A^{-1}$  (ver, p.ej., Peña (2002), p.35), entonces

$$\hat{\phi}_0 = M_t^{(2)} \left( M_t^{(1)} + M_t^{(2)} \right)^{-1} M_t^{(1)} \left( Y_{t-1}'^{(1)} Y_t^{(1)} + Y_{t-1}'^{(2)} Y_t^{(2)} \right),$$

por lo que

$$\begin{aligned}
 \hat{\phi}_0 &= M_t^{(2)} \left( M_t^{(1)} + M_t^{(2)} \right)^{-1} M_t^{(1)} X_t'^{(1)} Y_t^{(1)} + \dots \\
 &\quad M_t^{(1)} \left( M_t^{(1)} + M_t^{(2)} \right)^{-1} M_t^{(2)} X_t'^{(2)} Y_t^{(2)} \\
 &= M_t^{(2)} \left( M_t^{(1)} + M_t^{(2)} \right)^{-1} \hat{\phi} + M_t^{(1)} \left( M_t^{(1)} + M_t^{(2)} \right)^{-1} (\widehat{\phi + \delta}) \\
 &= \hat{\omega} \hat{\phi} + (1 - \hat{\omega}) (\widehat{\phi + \delta}).
 \end{aligned}$$

Así pues  $\hat{\phi}_0$  estará entre  $\hat{\phi}$  y  $\widehat{\phi + \delta}$ . □

### 3.3.2 Justificación del factor de olvido

Una vez estimado el valor inicial  $\hat{\phi}_0$  de la recursividad, se necesita que la secuencia de estimaciones  $\hat{\phi}_{\pi_t}$  este tan cerca de los valores reales como sea posible. Esto se consigue empleando un factor de olvido variable en (3.10).

Si se asume que el proceso  $Y_t$  sigue un modelo SETAR (3.25), es claro ver que si  $Y_t$  es ordenado respecto a la variable umbral  $Y_{t-d}$ , las primeras observaciones de la muestra ordenada, mientras se cumpla que  $Y_{t-d} \leq r$ , siguen un modelo autorregresivo  $Y_{\pi_t} = \phi X_{\pi_t} + e_t$ . En ese caso, se necesita que la secuencia de estimaciones  $\hat{\phi}_{\pi_t}$  se mueva desde  $\hat{\phi}_0$  hacia  $\phi$  lo más rápidamente posible, lo que implica el uso de un factor de olvido pequeño. Sin embargo, una vez que  $\hat{\phi}_{\pi_t}$  esté cerca de  $\phi$ , se necesita que el factor de olvido aumente para así conseguir reducir la variabilidad de la estimación.

Cuando  $Y_{t-d} > r$ , el modelo que seguirá el proceso es  $Y_{\pi_t} = (\phi + \delta) X_{\pi_t} + e_t$ , por lo que de nuevo se necesita un factor de olvido pequeño para que permita a  $\hat{\phi}_{\pi_t}$  aproximarse a  $\phi + \delta$ . Finalmente, una vez que  $\hat{\phi}_{\pi_t}$  esté cerca de  $\phi + \delta$  se necesitará, de nuevo, que el factor de olvido se incremente para reducir la variabilidad. De esta manera mediante el uso de un factor de olvido variable adecuado se puede obtener una secuencia de estimaciones  $\hat{\phi}_{\pi_t}$  con un codo alrededor del valor del umbral  $r$ .

Una vez justificado porqué es necesario emplear un factor de olvido variable, faltará justificar qué factor de olvido emplear. Para ello se repasan los problemas básicos que pueden surgir al emplear un método de estimación adaptativo (ver, p.ej., Rao Sri-pada and Grant Fisher 1987). El primero de los problemas habituales es que una matriz de ganancias  $M_t^{-1}$  con valores elevados en (3.8) provocará una alta variabilidad en la secuencia de estimaciones  $\hat{\phi}_{\pi_t}$ . Como resultado de ello, falsos cambios estructurales podrían ser detectados. Este problema puede aparecer cuando las observaciones entrantes al sistema recursivo  $X_{\pi_t}$  son iguales durante varios periodos de tiempo. El problema, sin embargo, puede ser limitado empleando un factor de olvido variable (3.17), ya que en estos casos el factor de olvido tenderá a incrementarse.

El segundo problema es el opuesto al anteriormente descrito. Si  $M_t^{-1}$  tiene valores muy pequeños, la cantidad que cambia la estimación recursiva (3.8) podría acabar siendo igual a 0. Cuando esto sucede se dice que la actualización recursiva se ha apagado. Esto podría suceder, por ejemplo, si el factor de olvido es demasiado elevado. Consecuentemente, un posible cambio estructural existente no sería detectado. Por ello, para que nuestro procedimiento funcione correctamente se necesita un factor de olvido variable que sea sensible a cambios en los parámetros del modelo. El factor de olvido (3.13) propuesto por Landau et al. (1998) no es una buena elección para nuestro objetivo, puesto que está relacionado con la influencia de las observaciones nuevas. Dado que los datos están ordenados de acuerdo a los valores crecientes de  $Y_{t-d}$ , las observaciones nuevas serán similares a las antiguas. Por ello, la medida de la influencia no nos ayudará a detectar cambio estructural alguno. Por el contrario, el factor de olvido propuesto en Sánchez (2006a) es especialmente conveniente para nuestra tarea. Este factor de olvido está basado en una representación recursiva de la distancia de Cook, permitiendo que el factor de olvido sea sensible a cambios en los parámetros del modelo.

### 3.3.3 Intervalos de confianza de $\phi_{s_t}$

El problema que nos surge ahora es cómo se puede evaluar si un codo observado en la secuencia de estimaciones  $\hat{\phi}_{\pi_t}$  es significativo, en el sentido de que realmente dicho codo ha sido provocado por un cambio de régimen en un modelo SETAR. Con este objetivo, la trayectoria observada  $\hat{\phi}_{\pi_t}$  es comparada con las trayectorias esperadas bajo el supuesto de que no hay ningún cambio de régimen. Esta comparación puede ser empleada para construir una región de confianza para la familia de trayectorias  $\hat{\phi}_{s_t}$  cuando el modelo es lineal. El problema es que la construcción de la región de confianza puede ser compleja. Un posible camino para construir la región es calculando las trayectorias correspondientes a todos los elementos  $s_i$  del conjunto  $S$ , es decir, calcular la estimación para todas las ordenaciones posibles. Esto nos daría un conjunto de trayectorias  $\hat{\phi}_{s_{i,t}}$ ,  $i = 1, \dots, n!$ . Entonces una región de confianza de cobertura  $100 \times (1 - \alpha)\%$  puede ser obtenida empleando los percentiles  $\alpha/2$  y  $1 - \alpha/2$  de dichas trayectorias. Por simplicidad, se emplean intervalos de confianza constantes a lo largo del tiempo. De esta manera las dos líneas que representan los percentiles de  $\hat{\phi}_{s_t}$  pueden ser interpretadas como una región de confianza dentro de la cual la trayectoria  $\hat{\phi}_{\pi_t}$  tenderá a permanecer bajo el supuesto de linealidad en los datos.

Un problema con el cálculo de los límites de confianza es que habitualmente no será factible calcular las  $n!$  trayectorias. La alternativa de calcular solo una muestra aleatoria de elementos de  $S$  es también costosa computacionalmente. Por ello, se propone emplear intervalos de confianza de  $\phi_{s_t}$  basados en una aproximación asintótica. Bajo

el supuesto de linealidad, las trayectorias  $\hat{\phi}_{s_t}$  y  $\hat{\phi}_{\pi_t}$  pueden ser tomadas como muestras aleatorias provenientes de la misma población. Entonces, para muestras grandes,  $\text{ECM}(\hat{\phi}_{s_{i,n}}) \approx \text{ECM}(\hat{\phi}_{\pi_n})$ ,  $\forall s_i \in S$ . Si el tamaño muestral  $n$  suficientemente grande, el intervalo de confianza asintótico  $\phi_{s_t}$ , bajo el supuesto de linealidad, puede ser aproximado por

$$\hat{\phi}_0 \pm Z_{1-\alpha/2} \sqrt{\hat{\sigma}_n^2 M_n^{-1}}. \quad (3.26)$$

Entonces, si la trayectoria  $\hat{\phi}_{\pi_t}$  tiene un codo fuera de los límites (3.26), se concluirá que hay un cambio de régimen y que el codo identifica el valor del umbral  $r$ .

### 3.3.4 Rendimiento en muestras finitas de los intervalos asintóticos

En esta sección se emplean experimentos de Monte Carlo para evaluar la validez de los intervalos asintóticos (3.26) como aproximación de los intervalos obtenidos con las trayectorias de estimaciones basadas en elementos aleatorios de  $S$ . Para ello se generan datos provenientes del modelo

$$Y_t = (-0.6 + \delta I_{(Y_{t-1} > 1)}) Y_{t-1} + e_t, \quad (3.27)$$

donde  $\delta$  tomará valores 0, en cuyo caso estaremos ante un modelo lineal AR(1), o 0.8, en cuyo caso estaremos ante un modelo SETAR(2;1,1). Además para cada uno de los  $\delta$  se emplea un tamaño muestral  $n = 150$  y  $n = 500$ . La secuencia  $e_t$  será un proceso de ruido blanco gaussiano tal que  $e_t \sim N(0, 3)$ .

Dado el modelo (3.27), si la secuencia de estimaciones recursivas  $\hat{\phi}_{\pi_t}$  tiende a salir fuera de los límites de confianza, se concluye que el modelo adecuado para el proceso  $Y_t$  es un SETAR. Además, el valor de  $Y_{t-1}$  correspondiente al valor de la secuencia  $\hat{\phi}_{\pi_t}$  más alejado de los límites de confianza será la estimación del valor del umbral  $r$ . La figura 3.1 muestra un ejemplo del funcionamiento de la herramienta ARLS sobre una réplica del modelo (3.27) para diferentes valores de  $\delta$  y  $n$ . En ella se puede ver la trayectoria de las estimaciones  $\hat{\phi}_{\pi_t}$  junto con varias líneas. La línea central continua es el estimador de mínimos cuadrados ordinarios  $\hat{\phi}_0$  usado como valor inicial de la recursión. La línea central discontinua es la media de 5000 trayectorias  $\hat{\phi}_{s_{i,t}}$  aleatorias muestreadas del conjunto de ordenaciones  $S$ . Los límites continuos son calculados mediante (3.26), mientras que los límites discontinuos muestran, para cada  $t$ , los percentiles empíricos 2.5 y 97.5 de las 5000 trayectorias aleatorias de  $\hat{\phi}_{s_{i,t}}$ .

Cuando  $\delta = 0$ , es decir, cuando el proceso es lineal, la figura 3.1 muestra que la secuencia de estimaciones  $\hat{\phi}_{\pi_t}$  permanece siempre dentro de los límites. Por el contrario, cuando  $\delta = 0.8$  la secuencia  $\hat{\phi}_{\pi_t}$  permanece fuera de los límites durante varios periodos de tiempo, mostrando un cambio estructural en los parámetros. En este caso, además

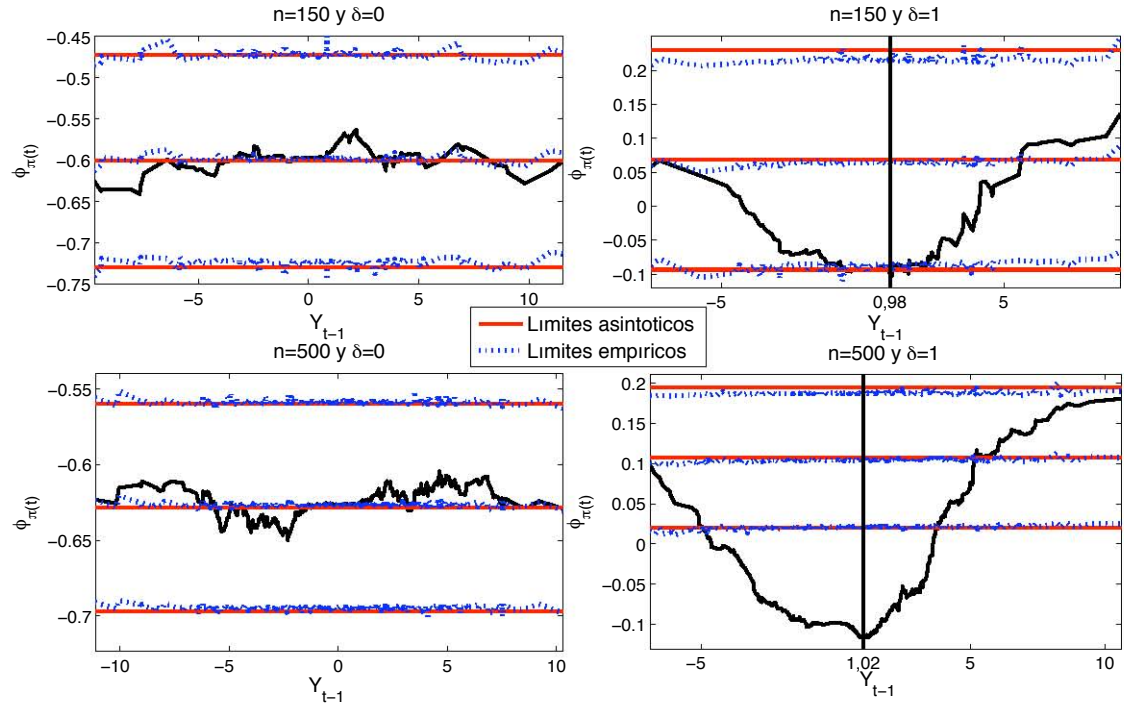


Figura 3.1: Ejemplo de funcionamiento del procedimiento ARLS

vemos como el valor más alejado de los límites está localizado alrededor de  $Y_{t-1} \approx 1$ , el cuál es el verdadero valor del umbral en el modelo generador de los datos (3.27).

En la figura 3.1 también podemos ver como los intervalos asintóticos parecen una buena aproximación de los intervalos empíricos calculados con 5000 ordenaciones aleatorias. Para validar este resultado, el experimento es repetido 1000 veces para cada valor de  $n$  y de  $\delta$ . Varias medidas son empleadas para comparar los intervalos de confianza. Primero se define una medida de la diferencia relativa entre los límites asintóticos y empíricos. Esta medida es denotada como  $B^l$  y se calcula mediante

$$B^l = \frac{1}{R} \sum_{k=1}^R \frac{\frac{1}{n} \sum_{i=1}^n l_a^{(k)} - l_{e_i}^{(k)}}{\left| U_a^{(k)} - L_a^{(k)} \right|}, \quad (3.28)$$

donde  $l$  tendrá valor  $U$  (límite superior) o  $L$  (límite inferior), y  $R$  es el número de replicaciones. Los subíndices  $a$  y  $e$  están referidos a los límites asintóticos y empíricos, respectivamente. Esta medida representa la distancia entre cada límite empírico y su aproximación asintótica y está estandarizada por la amplitud del intervalo. Esta estandarización es necesaria para dar una mejor idea del tamaño del sesgo respecto al

$\delta$	n=150					n=500				
	B <sup>L</sup>	B <sup>U</sup>	A	CV <sup>L</sup>	CV <sup>U</sup>	B <sup>L</sup>	B <sup>U</sup>	A	CV <sup>L</sup>	CV <sup>U</sup>
0	0.007	0.039	0.955	0.026	0.012	0.003	0.013	0.989	0.003	0.002
0.8	0.008	0.050	0.954	0.014	0.008	0.011	0.025	0.985	0.002	0.002

Tabla 3.1: Medidas para comparar los límites empíricos con los asintóticos empleando 1000 réplicas del modelo 3.27 para diferentes tamaños muestrales

tamaño del intervalo que se quiere estimar. La segunda medida está definida por

$$A = \frac{1}{R} \sum_{k=1}^R \frac{\frac{1}{n} \sum_{i=1}^n |U_{e_i}^{(k)} - L_{e_i}^{(k)}|}{|U_a^{(k)} - L_a^{(k)}|}, \quad (3.29)$$

y sirve para comparar la amplitud de los intervalos empíricos y asintóticos. El objetivo es conseguir que  $A \approx 1$ . Los límites empíricos han sido estimados empleando 5000 ordenaciones aleatorias provenientes de las  $n!$  posibles, por ello el coeficiente de variación de los límites empíricos ha sido también calculado para así comprobar su variabilidad.

La tabla 3.1 resume los resultados para diferentes valores de  $n$  y  $\delta$  en el modelo (3.27). Los coeficientes de variación son aproximadamente 0, lo que significa que 5000 ordenaciones aleatorias son suficientes para calcular los límites empíricos. Los valores de B<sup>L</sup> son muy pequeños, del orden de  $10^{-2}$  para cada valor de  $n$  y  $\delta$ . Finalmente, el valor medio  $A$  es próximo a 1. Para  $n = 150$  se obtiene  $A \approx 0.95$ , lo que significa que la amplitud de los límites asintóticos son aproximadamente un 5% mayores que los empíricos, es decir, los límites asintóticos son ligeramente conservadores. En resumen, se puede concluir que los resultados de las simulaciones confirman que los intervalos de confianza asintóticos (3.26) son una buena aproximación de los intervalos empíricos.

### 3.3.5 Eficacia del método ARLS en muestras finitas

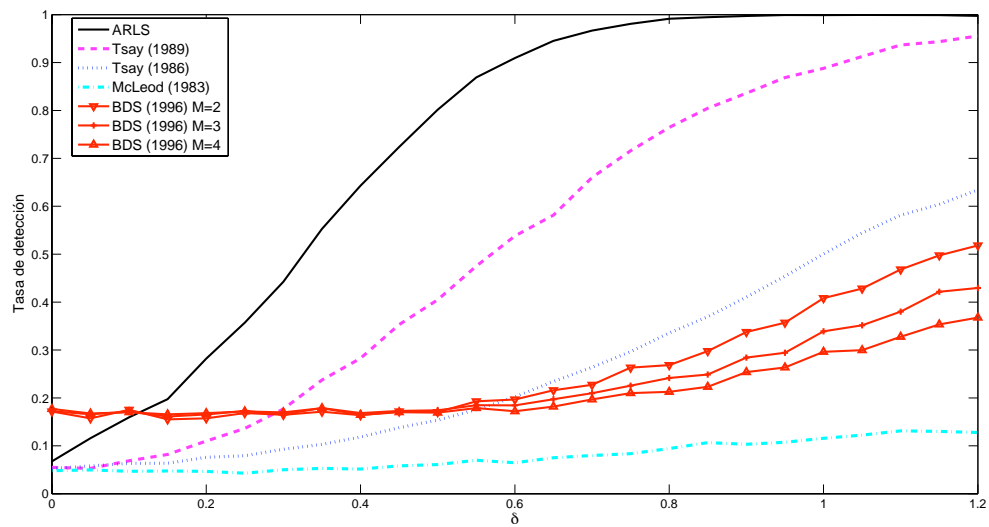
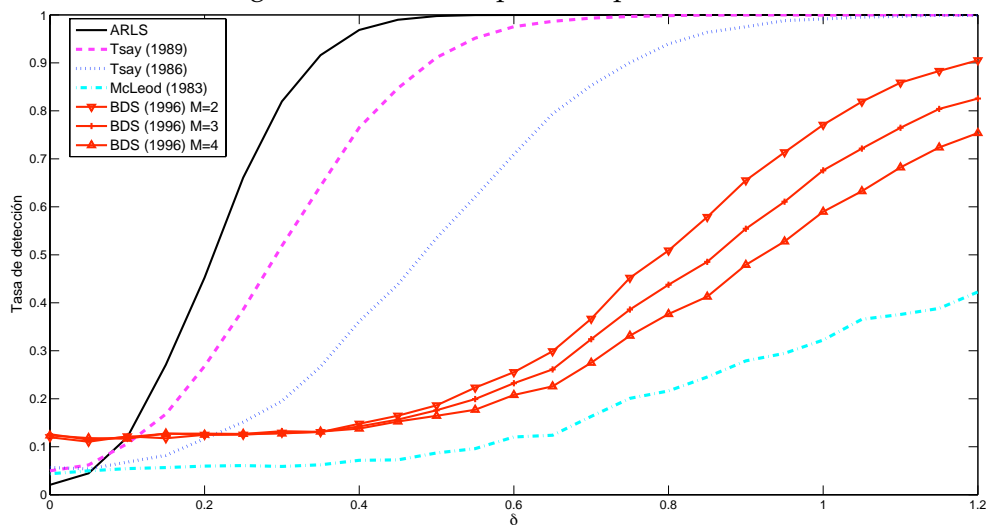
En esta sección se evalúa la eficiencia del método ARLS para detectar modelos SETAR empleando experimentos de Monte Carlo. Para ello el método ARLS propuesto es comparado con varios métodos existentes en la literatura previa. Los métodos empleados en la comparación son tanto métodos generales para detectar no linealidad como métodos específicos para modelos SETAR. Entre los métodos generales, están incluidos en la evaluación la propuesta de Tsay (1986), la de McLeod y Li (1983) y el contraste BDS propuesto por Brock, Dechert and Scheinkman (1996). Los contrastes para modelos SETAR incluidos en la competición son las propuestas de Tsay (1989) y Hansen (1996, 1999). El nivel de significación empleado en todos los contrastes es  $\alpha = 0.05$ , mientras que el método ARLS está basado en los intervalos de confianza asintóticos del 95%.

Métodos	$\delta$				
	0	0.25	0.5	0.75	1
McLeod y Li (1983)	0.06	0.04	0.07	0.13	0.22
Tsay (1986)	0.02	0.14	0.29	0.59	0.89
Tsay (1989)	0.03	0.33	0.60	0.94	1.00
BDS2	0.14	0.13	0.15	0.32	0.56
BDS3	0.12	0.15	0.16	0.24	0.44
BDS4	0.15	0.10	0.11	0.22	0.35
Hansen Bootstrap Homoscedástico	0.04	0.06	0.15	0.36	0.76
Hansen Bootstrap Heterocedástico	0.03	0.04	0.12	0.27	0.68
Hansen Asintótico Homocedástico	0.19	0.29	0.40	0.62	0.90
Hansen Asintótico Heterocedástico	0.01	0.02	0.06	0.21	0.56
ARLS	0.06	0.41	0.96	1.00	1.00

Tabla 3.2: Tasa de detección para diferentes contrastes en datos generados empleando el modelo (3.27) para diferentes valores de  $\delta$

Para realizar el primer experimento se generan realizaciones provenientes del modelo (3.27) con  $\delta = (0, 0.25, 0.5, 0.75, 1)$ . En este experimento, los contrastes de Hansen (1999) son implementados empleando los códigos de Matlab obtenidos en la página web del autor, por esta razón, el tamaño muestral es ajustado a  $n = 289$ , puesto que esta es la dimensión de la serie de datos que emplea Hansen en el ejemplo que resuelve en sus códigos. Debido al coste computacional de los contrastes de Hansen, el número de replicaciones de este experimento es restringido a 100. La tabla 3.2 muestra la tasa de detección para cada contraste. Se puede ver que los mejores resultados son obtenidos por el método ARLS.

El experimento es repetido empleando 5000 replicaciones del modelo (3.27) para valores alternativos de  $\delta \in [0, 1.2]$  y  $n = (150, 500)$ , pero sin los contrastes de Hansen, ya que son demasiado costosos computacionalmente. Las figuras 3.2 y 3.3 muestran las tasas de detección obtenidas. Los resultados confirman que el método ARLS propuesto es el más eficiente. También puede verse como los contrastes de no linealidad generales tienen una potencia baja para detectar no linealidad provocada por los umbrales.

Figura 3.2: Curva de potencia para  $n=150$ Figura 3.3: Curva de potencia para  $n=500$



régimen 1			régimen 2			r	n=100 y 1000 réplicas			n=200 y 500 réplicas		
$\phi_0^{(1)}$	$\phi_1^{(1)}$	$\sigma^{(1)}$	$\phi_0^{(2)}$	$\phi_1^{(2)}$	$\sigma^{(2)}$		Hansen	Tsay	ARLS	Hansen	Tsay	ARLS
0	0.3	1	0	0.3	1	-	0.045	0.040	0.066	0.054	0.049	0.063
-0.75	0.3	1	0	0.3	2	-0.76	0.193	0.040	0.374	0.312	0.075	0.338
-1.25	0.3	1	0	0.3	1	-0.97	0.741	0.212	0.744	0.982	0.376	0.990
-1.25	0.3	1	0	0.3	2	-1.25	0.391	0.126	0.480	0.716	0.239	0.793
0	-0.3	1	0	0.3	2	0.25	0.347	0.218	0.564	0.718	0.493	0.856
0	-0.7	1	0	0.3	1	0.34	0.826	0.734	0.902	0.984	0.969	1
0	-0.7	1	0	0.3	2	0.49	0.865	0.694	0.944	0.998	0.964	1
-1.25	-0.7	1	0	0.3	1	-0.2	0.913	0.858	0.980	0.996	0.982	0.998
-1.25	-0.7	1	0	0.3	2	-0.1	0.876	0.694	0.996	1	0.974	1
-1.25	-0.7	2	0	0.3	1	0.15	0.958	0.946	0.980	0.998	0.999	1

Tabla 3.3: Tasa de detección obtenida empleando ARLS y el contraste de Tsay (1989) y el mejor de los resultados obtenido por los contrastes de Hansen (1996)

Se realiza un segundo experimento basado en el realizado en Clements et al. (2003). El experimento de Monte Carlo está basado en datos simulados provenientes de 10 modelos con diferentes características. Para  $n = 100$  los autores fijaron como número de replicaciones 1000, y para  $n = 200$ , el número de replicaciones es 500. La tabla 3.3 muestra para cada modelo los mejores resultados mostrados por Clements et al. (2003) obtenidos empleando alguno de los contrastes propuestos en Hansen (1996), y nuestros resultados para el método ARLS y el contraste de Tsay (1989). La mejor tasa de detección es siempre obtenida por el método ARLS.

Como último experimento se evalúa la eficiencia del método ARLS para estimar los valores de los umbrales. Para ello se generan 5000 replicaciones del modelo (3.27) para valores de  $\delta \in [0, 1]$  y  $n = 500$ . Cuando el método ARLS detecta un modelo SETAR, se estima un umbral  $\hat{r}$ . Entonces, se calcula el histograma de estos valores estimados  $\hat{r}$  para cada  $\delta$ . La figura 3.4 muestra los resultados, mostrando que el histograma es apuntado alrededor de 1, que es el verdadero valor del umbral.

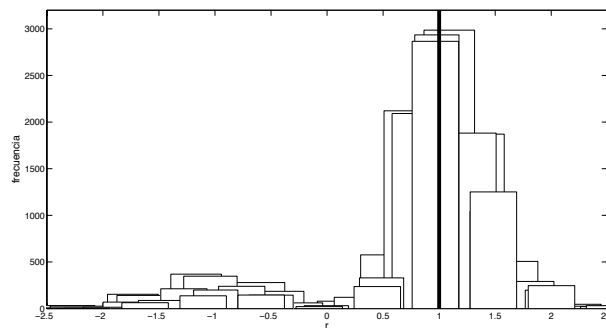


Figura 3.4: Histograma de los umbrales estimados para diferentes valores de  $\delta$

### 3.4 Procedimiento automático, Aut-ARLS

En la sección anterior, se ha mostrado un procedimiento que funciona de manera más eficiente que los contrastes previos existentes para detectar modelos SETAR. El principal problema de estos métodos es que sólo pueden ser empleados para detectar modelos sencillos con pocos regímenes. Por ello, empleando los resultados previos, se desarrolla un procedimiento que detecte y modelice modelos SETAR de manera automática. Dicho procedimiento ha sido llamado Aut-ARLS y nos permitirá encontrar modelos más complejos si los hubiera. El procedimiento automático funciona de la siguiente manera.

- Paso 1. Seleccionar el orden autorregresivo  $p$  empleando para ello la función de autocorrelación parcial y el criterio de información de Akaike. Seleccionar un valor máximo  $d_{max}$  para el parámetro *delay*.
- Paso 2. Ajustar un TV-AAR( $p$ ) ordenado en orden ascendente de acuerdo a  $Y_{t-d}$  para cada valor de  $d = 1, \dots, d_{max}$ , usando el factor de olvido (3.17). Calcular el intervalo de confianza asintótico empleando la expresión (3.26) para cada parámetro del vector  $\hat{\phi}$ . Si hay valores de la secuencia de estimaciones  $\hat{\phi}_{j\pi_t}, j = 1, \dots, p$ , que salen fuera de los límites asintóticos, los valores correspondientes de  $Y_{t-d}$  son escogidos como un conjunto de posibles umbrales  $r^{asc} = \left\{ Y_{t-d} \mid |\hat{\phi}_{\pi_t} - \hat{\phi}_0| > Z_{1-\alpha/2} \sqrt{\hat{\sigma}_n^2 M_n^{-1}} \right\}$ . Además, las distancias correspondientes a los límites asintóticos son también guardados en  $h^{asc} = \left\{ |\hat{\phi}_{\pi_t} - \hat{\phi}_0| \mid Y_{t-d} \in r^{asc} \right\}$ .
- Paso 3. Repetir el Paso 2 con  $Y_t$  ordenado respecto al orden descendente de  $Y_{t-d}$ . Denotaremos como  $\eta_t$  a la  $t$ -ésima posición del elemento de  $S$  correspondiente a la ordenación de los casos en el orden descendente de la variable  $Y_{t-d}$ . Denotaremos la secuencia de las estimaciones adaptativas como  $\hat{\phi}_{j\eta_t}, j = 1, \dots, p$ . El uso de secuencias de estimación basadas tanto en orden ascendente como descendente de  $Y_{t-d}$ , puede ser útil cuando el umbral está próximo a un extremo de la variable umbral, eliminando la posibilidad de que el umbral quede enmascarado por la variabilidad de las estimaciones iniciales. Como en el paso anterior, obtenemos un conjunto de posibles umbrales  $r^{desc} = \left\{ Y_{t-d} \mid |\hat{\phi}_{\eta_t} - \hat{\phi}_0| > Z_{1-\alpha/2} \sqrt{\hat{\sigma}_n^2 M_n^{-1}} \right\}$  y sus respectivas distancias  $h^{desc} = \left\{ |\hat{\phi}_{\eta_t} - \hat{\phi}_0| \mid Y_{t-d} \in r^{desc} \right\}$ .
- Paso 4. Estimar el valor del umbral  $r$  para cada valor del *delay*  $d$  entre los candidatos  $r^{asc}$  y  $r^{desc}$ . Para ello los conjuntos  $r^{asc}$  y  $r^{desc}$ , y sus respectivas distancias  $h^{asc}$  y  $h^{desc}$ , son combinadas. Si hay candidatos que repiten en ambos conjuntos  $r^{asc}$  y  $r^{desc}$ , sus distancias correspondientes en  $h^{asc}$  y  $h^{desc}$  son sumadas. Entonces, la estimación del umbral  $r$  es el candidato con una distancia más grande.

- Paso 5. Para cada valor de  $d$ , ajustar un modelo SETAR empleando los valores estimados de  $r$ , si los hubiera. El orden autorregresivo en cada régimen es seleccionando empleando el criterio de selección de modelos propuesto por Galeano y Peña (2007).
- Paso 6. Repetir los Pasos 2-5 sobre las series residuales obtenidas en el Paso 5, cuando  $Y_{t-d} \leq r$  y  $Y_{t-d} > r$ . Ya que la varianza puede ser diferente en cada régimen, este paso será realizado separando los residuos de cada régimen y analizando cada régimen por separado. El procedimiento es repetido hasta que no se encuentre ningún umbral adicional.
- Paso 7. Seleccionar el modelo SETAR final. Si varios modelos SETAR son detectados para diferentes valores de  $d$ , el modelo SETAR final es seleccionado usando el criterio de Galeano y Peña (2007).

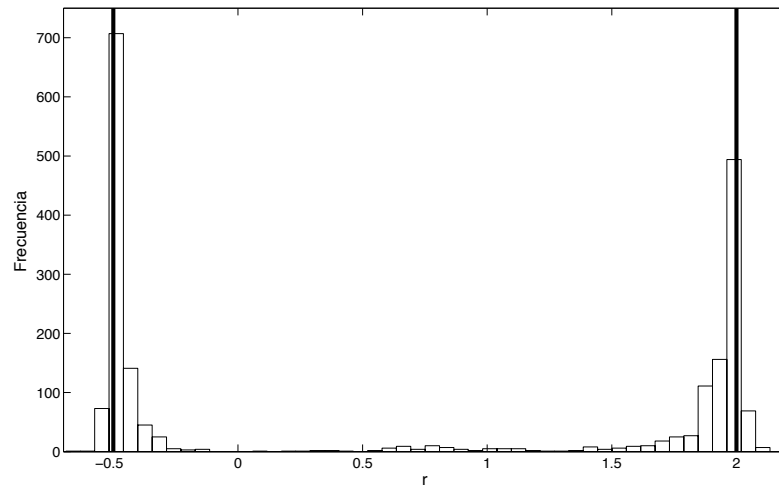


Figura 3.5: Histograma de los umbrales estimados empleando Aut-ARLS sobre 1000 réplicas del modelo (3.30)

La eficiencia del método Aut-ARLS es evaluada mediante simulaciones. El primer experimento consiste en simular 1000 replicas del modelo SETAR(3;[2,3,1],2),

$$Y_t = \begin{cases} -0.7Y_{t-1} + 0.1Y_{t-2} + e_t^{(1)} & , \text{ si } Y_{t-2} \leq -0.5; \\ 0.2Y_{t-1} + 0.6Y_{t-2} - 0.3Y_{t-3} + e_t^{(2)} & , \text{ si } -0.5 < Y_{t-2} \leq 2; \\ 0.8Y_{t-1} + e_t^{(3)} & , \text{ si } Y_{t-2} > 2. \end{cases} \quad (3.30)$$

donde la secuencia  $e_t^{(j)}$  será un proceso de ruido blanco gaussiano con distribución  $N(0, \sigma^{(j)})$ , siendo  $\sigma^{(j)} = (3, 2, 5)$ . El tamaño muestral es  $n = 500$ . El método Aut-ARLS es aplicado a los datos simulados, obteniendo para cada réplica el parámetro  $d$

y el umbral  $r$ . La tasa de detección de un modelo SETAR empleando Aut-ARLS es 100%. Además, el porcentaje de replicaciones en las cuales Aut-ARLS detecta la variable umbral correcta es también 100%. Finalmente, calculamos la media del número de umbrales detectados por Aut-ARLS, obteniendo 2.02. La figura 3.5 muestra el histograma de los umbrales estimados para cada replicación, en el que podemos ver dos picos entorno a  $-0.5$  y  $2$ , los cuales son los verdaderos valores de los umbrales.

En el segundo experimento el modelo que genera los datos es

$$Y_t = (-0.6 + \delta_1 I_{(-0.5 < Y_{t-2} \leq 0.5)} + \delta_2 I_{(0.5 < Y_{t-2})}) Y_{t-1} + e_t, \quad (3.31)$$

donde  $(\delta_1, \delta_2) \in [0, 1]$  y  $n = 500$ . La secuencia  $e_t$  será un proceso de ruido blanco gaussiano con distribución  $N(0, 3)$ . El experimento consiste en la simulación de 500 réplicas para valores alternativos de  $\delta_1$  y  $\delta_2$ . La figura 3.6 muestra la tasa de detección mientras que la figura 3.7 da el porcentaje de replicaciones en las cuales Aut-ARLS detecta la verdadera variable umbral. De los gráficos se deduce que es claro que Aut-ARLS funciona correctamente.

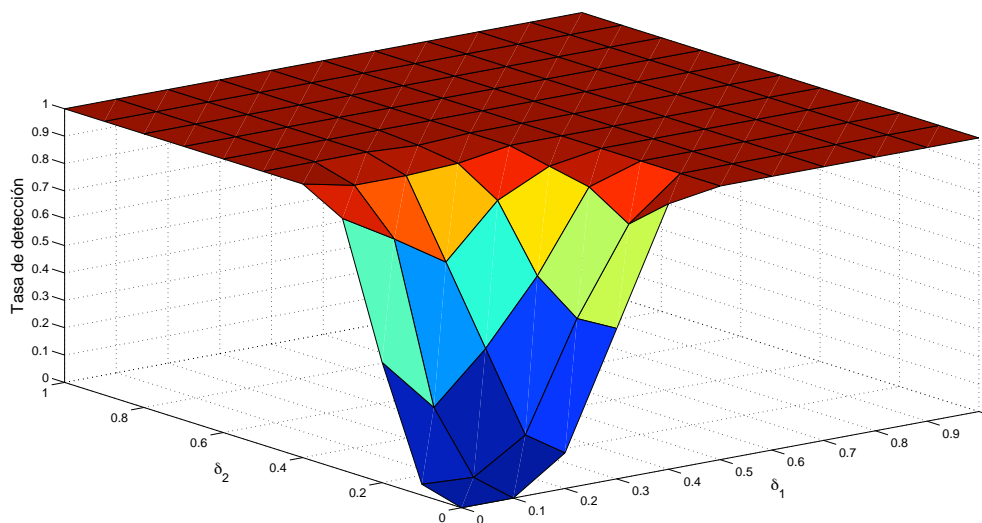


Figura 3.6: Tasa de detección para diferentes valores de  $\delta_1$  y  $\delta_2$  obtenida empleando Aut-ARLS

### 3.5 Aplicación en datos reales clásicos

En esta sección el procedimiento automático Aut-ARLS es aplicado en ejemplos de datos reales. Para ello se han empleado los datos de los linces canadienses, de las manchas solares y del producto nacional bruto de Estados Unidos, los cuales han sido

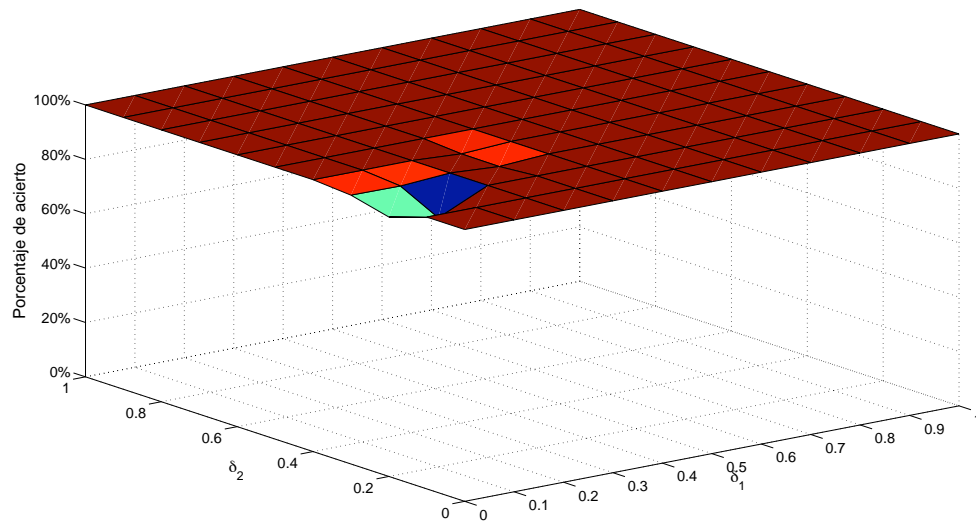


Figura 3.7: Porcentaje de acierto de Aut-ARLS

ampliamente estudiados en la literatura.

### 3.5.1 Lince canadienses

La serie consta del número de lince capturados anualmente en el distrito *Mackenzie River* en el noroeste de Canadá durante los años 1821 y 1934. Esta serie de datos ha sido ampliamente analizada en multitud de estudios en el campo de las series temporales. Los datos pueden encontrarse en multitud de páginas web y libros (por ejemplo, Tong 1990). En la figura 3.8 podemos ver la serie original transformada por logaritmos. Dicha transformación se hace habitualmente al trabajar con esta serie desde el trabajo de Moran (1953).

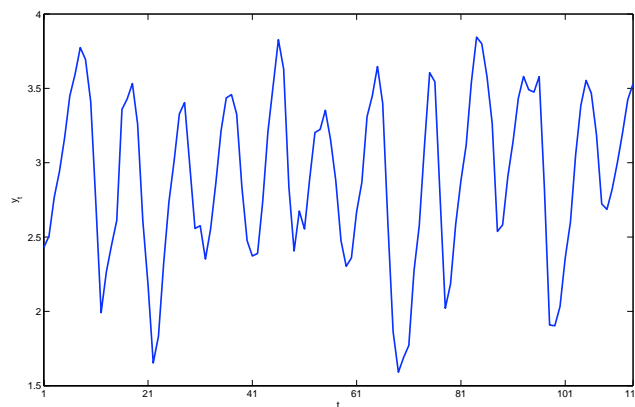


Figura 3.8: Serie de los lince canadienses transformada con logaritmos

Propuesta	Delay	Umbral	Ordenes SETAR	AIC	BIC
Tsay (1989)	2	(2.373, 4.154)	(1, 7, 2)	-347.7	-322.4
Tong (1990)	2	3.116	(7, 2)	-337.6	-315.2
Aut-ARLS	2	3.2639	(3, 2)	-353.1	-339.0

Tabla 3.4: Modelos propuestos por Tsay, Tong y Aut-ARLS con sus respectivos AIC y BIC

Aplicamos pues la herramienta Aut-ARLS a los datos transformados por logaritmos, seleccionando como orden autorregresivo del modelo TV-AAR (3.2) a estimar  $p = 2$  y como variable umbral el segundo retardo  $Y_{t-2}$ . La figura 3.9 muestra que el umbral detectado es 3.2639.

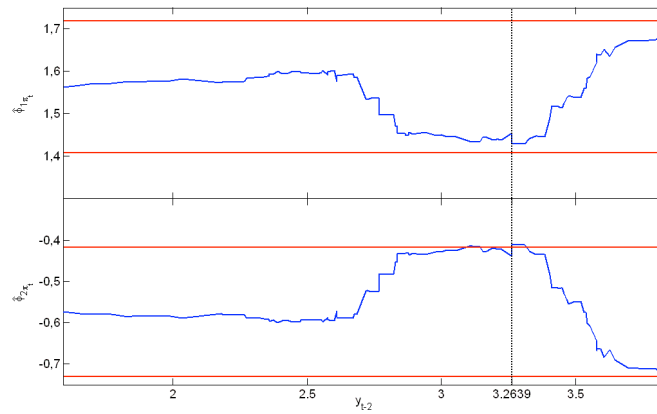


Figura 3.9: Detección del umbral en los datos de los lince canadienses empleando Aut-ARLS

La tabla 3.4 resume el modelo propuesto por Aut-ARLS y los modelos propuestos en los estudios previos de Tong (1990) y Tsay (1989). Los valores mínimos correspondientes a los criterios AIC y BIC son encontrados para el modelo propuesto por Aut-ARLS.

### 3.5.2 Manchas solares

Esta popular serie de datos consiste en el número de manchas solares anuales observadas desde 1700 a 2008. Los datos están disponibles en la página web del *National Geophysical Data Center*. Para realizar la identificación del modelo SETAR han sido empleadas las observaciones pertenecientes al periodo de tiempo que abarca desde 1700 hasta 1920, debido a que estas fueron las observaciones que poseía Tong (1983). De esta manera nos garantizamos que no poseemos mayor información que en los estudios previos. La figura 3.10 muestra el número anual de manchas solares.

Si aplicamos Aut-ARLS, la herramienta selecciona como orden autorregresivo un

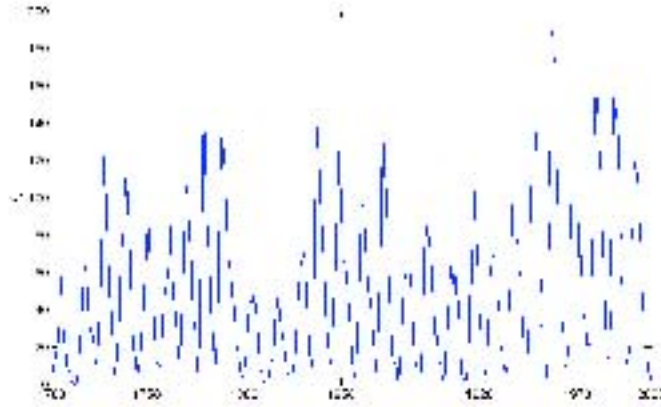


Figura 3.10: Serie de las manchas solares

$p = 3$  y detecta como variable umbral  $Y_{t-3}$ . La figura 3.11 muestra las secuencias de estimaciones recursivas  $\hat{\phi}_{\pi_t}$  del modelo TV-AAR(3) ordenado según  $Y_{t-3}$ . En particular, las figuras 3.11a y 3.11b muestran las estimaciones de las ordenaciones ascendentes y descendentes, respectivamente. En la ordenación ascendente, la observación 30.7 es el principal candidato a ser seleccionado como umbral, mientras que en la ordenación descendente existen varios candidatos. Aut-ARLS selecciona como valor del umbral 30.7, empleando los criterios explicados anteriormente.

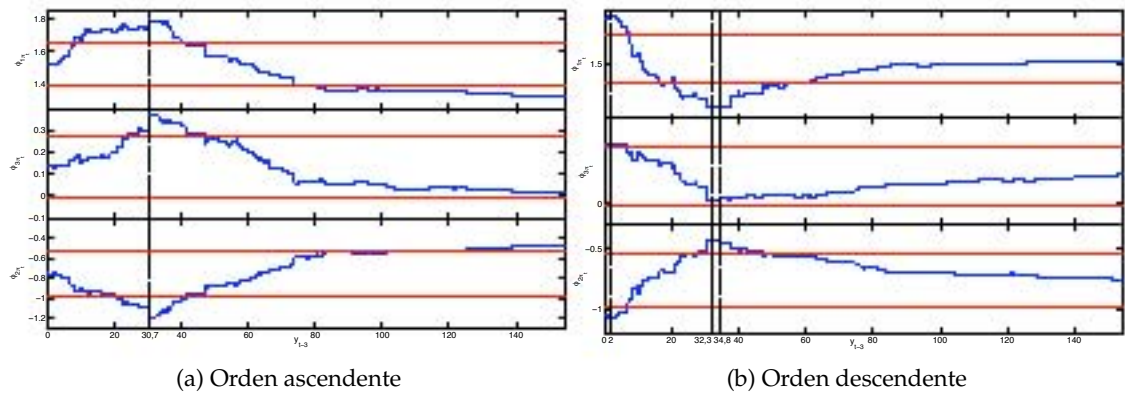


Figura 3.11: Detección del umbral empleando Aut-ARLS para los datos de las manchas solares

Una vez que el umbral ha sido seleccionado, un modelo SETAR puede ser estimado. Aut-ARLS busca más posibles umbrales en los residuos de cada régimen, pero ningún umbral adicional fue encontrado. Para comparar el modelo seleccionado por Aut-ARLS hemos empleado los modelos SETAR propuestos por Tsay (1989) y Tong (1983) para obtener predicciones fuera de la muestra del índice de manchas solares durante el pe-

Propuestas	Delay	Umbral	Órdenes SETAR	AIC <sub>1700-1920</sub>	EMA <sub>1921-2008</sub>		RECM <sub>1921-2008</sub>	
					h=1	h=2	h=1	h=2
Tong (1983)	3	36.6	(3,11)	1083.8	12.31	12.42	16.59	16.69
Tsay (1989)	2	(34.8, 70.9)	(11,10,10)	1064.1	11.74	11.87	15.52	15.61
Aut-ARLS	3	30.7	(7,11)	1045.9	11.37	11.47	15.33	15.42

Tabla 3.5: Modelos propuestos por Tsay, Tong y Aut-ARLS con sus respectivos AIC, error medio absoluto y raíz del error cuadrático medio.

riodo de tiempo 1921-2008. La tabla 3.5 resume los resultados para los modelos propuestos. Los valores del AIC fueron hallados por Tsay (1989) con los datos del periodo de tiempo 1700-1920, mientras que para las predicciones fuera de la muestra son calculados la raíz del error cuadrático medio (RECM) y el error medio absoluto (EMA) para horizontes de predicción  $h = 1, 2$ . El modelo propuesto por Aut-ARLS tiene menor valor del AIC, el RECM y el EMA que las propuestas de Tong y Tsay.

### 3.5.3 Producto nacional bruto EEUU

Una de las principales características de los modelos SETAR es que permiten la modelización de ciclos asimétricos en los datos, por ello su utilidad en el campo de la economía ha sido ampliamente investigada. Entre las distintas series económicas analizadas se encuentra la del Producto Nacional Bruto (PNB) cuatrimestral de Estados Unidos. El análisis de los datos del PNB se realiza sobre la tasa de crecimiento, calculada mediante

$$Y_t = \log(X_t) - \log(X_{t-1}),$$

donde  $X_t$  es la serie del PNB desestacionalizada. En la figura 3.12 vemos la tasa de crecimiento del PNB desde el primer cuatrimestre de 1947 al segundo de 2003.

Tsay y Tiao (1994) propusieron el modelo SETAR con cuatro regímenes definido por

$$Y_t = \begin{cases} -0.015 - 1.076Y_{t-1} + e_t^{(1)} & , \text{ si } Y_{t-1} \leq Y_{t-2} \leq 0 \\ -0.006 + 0.63Y_{t-1} - 0.756Y_{t-2} + e_t^{(2)} & , \text{ si } Y_{t-1} > Y_{t-2} \text{ y } Y_{t-2} \leq 0 \\ 0.006 + 0.438Y_{t-1} + e_t^{(3)} & , \text{ si } Y_{t-1} \leq Y_{t-2} \text{ y } Y_{t-2} > 0 \\ 0.004 + 0.443Y_{t-1} + e_t^{(4)} & , \text{ si } Y_{t-1} > Y_{t-2} > 0 \end{cases}, \quad (3.32)$$

que permite explicar la evolución del PNB en términos de periodos de recesión y expansión económica. En su trabajo Tiao y Tsay (1994) muestran cómo en determinados periodos el modelo SETAR predice mejor que un modelo lineal. Peña y Rodríguez (2005b) analizaron la serie empleando diferentes tipos de modelos lineales con un modelo SETAR. Para comparar las diferentes propuestas realizaron predicciones fuera de



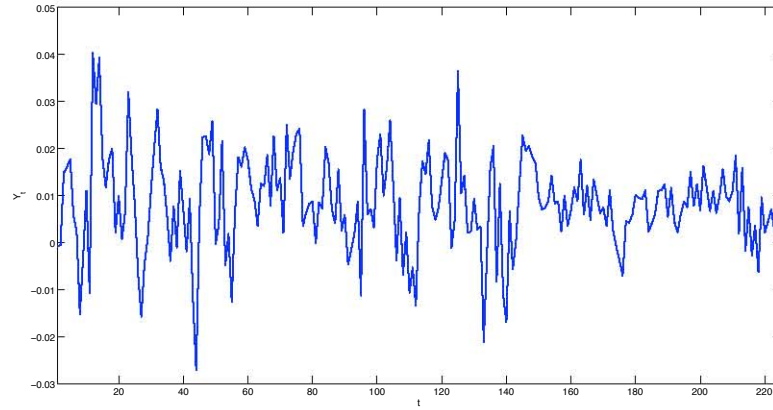


Figura 3.12: Tasa de crecimiento del producto nacional bruto de Estados Unidos desde 1947 a 2003.

la muestra empleando las 10 últimas observaciones. El modelo SETAR propuesto es

$$Y_t = \begin{cases} 0.0031 + 1.0345Y_{t-1} + e_t^{(1)} & , \text{ si } Y_{t-1} \leq 0.0084 \\ 0.0045 + 0.2340Y_{t-1} + e_t^{(2)} & , \text{ si } Y_{t-1} > 0.0084 \end{cases} \quad (3.33)$$

mientras que el mejor de los modelos lineales empleados es el proceso autorregresivo con dos atípicos aditivos, un cambio transitorio y un cambio de nivel siguiente

$$Y_t = 0.423Y_{t-1} + w_1I_t^{(04/49)} + w_2I_t^{(04/70)} + \frac{w_3}{1-\delta B}I_t^{(01/70)} + w_4S_t^{(02/78)} + e_t, \quad (3.34)$$

siendo los  $w_i$  el tamaño de los atípicos cuya influencia vendrá dada por la variable impulso  $I_t^{(k)} = 0, t \neq k; I_k^{(k)} = 1$ , y la variable escalón  $S_t^{(k)} = 1, t \geq k$ , y cero en caso contrario. Peña y Rodríguez (2005b) muestran que el modelo que obtiene menor error cuadrático medio es el autorregresivo con los atípicos ajustados (3.34), mejorando a los modelos SETAR propuestos.

Se aplica el procedimiento Aut-ARLS a la tasa de crecimiento del PNB para comprobar si se puede obtener un modelo SETAR mejor que las propuestas previas, encontrando el modelo SETAR con 4 regímenes siguiente

$$Y_t = \begin{cases} 0.0078 + 0.3927Y_{t-1} + e_t^{(1)} & , \text{ si } \Delta Y_{t-1} \leq 0.0075 \text{ y } Y_{t-2} \leq 0 \\ 0.0040 + 0.3034Y_{t-1} + 0.1741Y_{t-2} + e_t^{(2)} & , \text{ si } \Delta Y_{t-1} \leq 0.0075 \text{ y } Y_{t-2} > 0 \\ -0.0068 + 0.5398Y_{t-1} - 0.8094Y_{t-2} + e_t^{(3)} & , \text{ si } \Delta Y_{t-1} > 0.0075 \text{ y } Y_{t-2} \leq 0 \\ 0.0071 + 0.2638Y_{t-1} + e_t^{(4)} & , \text{ si } \Delta Y_{t-1} > 0.0075 \text{ y } Y_{t-2} > 0 \end{cases} \quad (3.35)$$

En la tabla 3.6 se muestra el error cuadrático medio de los diferentes modelos. Como vemos, el modelo SETAR propuesto por Aut-ARLS mejora a las anteriores propuestas

Propuestas	Modelo	ECM <sub>01/01-02/03</sub>
Tiao y Tsay	SETAR	$3.08 \cdot 10^{-5}$
Peña y Rodríguez	AR con atípicos	$1.67 \cdot 10^{-5}$
Peña y Rodríguez	SETAR	$4.66 \cdot 10^{-5}$
Aut-ARLS	SETAR	$1.33 \cdot 10^{-5}$

Tabla 3.6: Error cuadrático medio de los errores de predicción del PNB EE.UU..

existentes. Esto muestra que la eficiencia del modelo SETAR depende en gran manera de una correcta estimación, lo que no es sencillo de obtener.

### 3.6 Conclusiones y líneas futuras de investigación

La identificación de un modelo no lineal es una tarea en la mayoría de las ocasiones compleja. Por una parte está la tarea de identificar el fenómeno no lineal que se encuentra en los datos y por otra parte está la modelización de la estructura del modelo. En el caso de los modelos SETAR existen diferentes contrastes de linealidad que permiten comprobar si un modelo SETAR es una alternativa adecuada en nuestros datos, sin embargo no existía un procedimiento sencillo que permitiera identificar la estructura del modelo, lo que hace reducir su eficacia.

A lo largo de este capítulo mostramos diferentes resultados que nos permiten proponer un procedimiento automático que de manera sencilla y mediante una herramienta gráfica permite estimar la estructura del modelo SETAR a emplear. Los resultados mostrados, tanto mediante simulaciones como mediante datos reales, confirman que el procedimiento propuesto mejora a sus predecesores, tanto en eficacia como en coste computacional.

En el ejemplo del producto nacional bruto de Estados Unidos comprobamos que la estimación de los umbrales es clave para que el modelo SETAR sea eficiente, por ello una posible mejora futura del procedimiento es afinar al máximo posible la identificación de los umbrales.

## Capítulo 4

# Densidad predictiva basada en momentos condicionados y máxima entropía. Aplicación a la predicción de potencia eólica

El cálculo de predicciones puntuales junto con su incertidumbre en forma de intervalo es, en la mayoría de aplicaciones, insuficiente, especialmente cuando estemos asumiendo no linealidad en los datos. Por ello será necesario, además de la predicción puntual, obtener una estimación de la densidad condicionada de la variable en el futuro dado su comportamiento actual. A esta distribución condicionada se le conoce con el nombre de densidad predictiva.

### 4.1 Energía eólica

La estimación de la densidad predictiva dependerá de los datos con los que se está trabajando. Por ejemplo, si los residuos del modelo empleado son gaussianos, el problema de estimar la densidad predictiva quedará reducido a la estimación de la media y la varianza condicionales. Sin embargo, esto no será adecuado en otras ocasiones en las que, por ejemplo, se pueden encontrar densidades predictivas no gaussianas. El objetivo de este capítulo es mejorar la estimación de la densidad predictiva de las predicciones puntuales de potencia eólica. Es decir, para unas predicciones puntuales proporcionadas por cierto agente, queremos estimar su distribución condicionada.

Para analizar los diferentes métodos propuestos a lo largo del capítulo vamos a emplear datos de potencia eólica media producida horariamente en 6 parques eólicos dife-

rentes. Para cada uno de dichos parques eólicos tendremos la producción media horaria observada y las predicciones de dicha producción a diferentes horizontes. Desconocemos cómo han sido calculadas las predicciones, por lo que no tenemos ninguna información previa acerca de las propiedades de los predictores más allá de lo que digan los datos disponibles. La frecuencia de dichos datos es horaria.

#### 4.1.1 Análisis descriptivo de la distribución de la potencia generada

La principal característica de la potencia eólica es que es una variable acotada, entre 0 y la potencia nominal, es decir, el valor máximo que puede producir un aerogenerador o un parque eólico, dependiendo de los datos con los que estemos trabajando. En general, se trabaja estandarizando la potencia producida por la potencia nominal, con lo que los datos estarán acotados entre 0 y 1. Los datos que hemos empleado en este capítulo pertenecen a la potencia eólica horaria generada por diferentes parques eólicos. Para esas potencias medidas tenemos las predicciones puntuales elaboradas por cierto agente para diferentes horizontes de predicción. El objetivo es estimar la función de densidad predictiva a partir de la serie temporal de predicciones y los datos históricos observados, sin tener ninguna información acerca del método empleado en la elaboración de las predicciones. Esta situación es muy común en sectores como el eléctrico, donde la metodología empleada en la elaboración de las predicciones forma parte del *"know-how"* de los diferentes proveedores de predicciones, y se mantiene un alto grado de confidencialidad. Como vemos en la figura 4.1 la potencia eólica es una variable no estacionaria que puede cambiar de manera abrupta. En la figura pequeña en la que mostramos un tramo de la serie vemos cómo la serie alterna rápidos cambios de nivel.

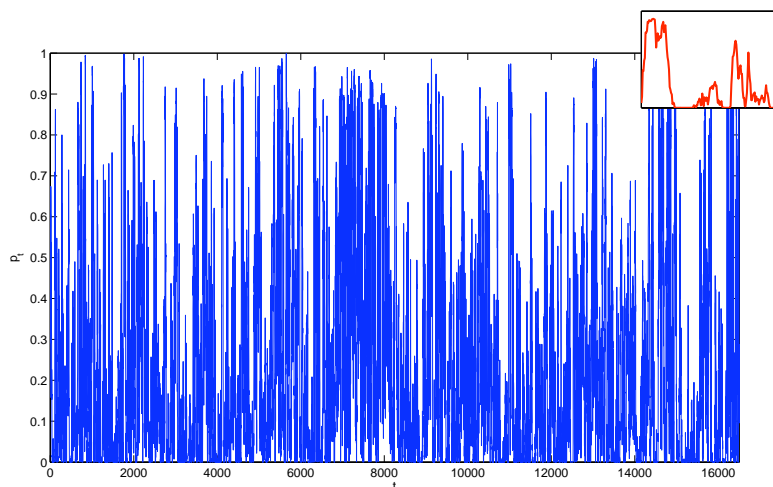


Figura 4.1: Potencia producida por un parque eólico, datos horarios

El objetivo del capítulo es proponer un procedimiento para estimar la densidad predictiva de las predicciones para diferentes horizontes de predicción. En la figura 4.2 vemos los gráficos de caja de la potencia medida para diferentes niveles de las predicciones. En dichas figuras vemos cómo existen asimetrías y varianzas que dependen del nivel de predicción independientemente del horizonte. Es decir, los errores de predicción tienen algunas características que pueden observarse a cualquier horizonte de predicción. Utilizaremos este tipo de patrones en el método propuesto.

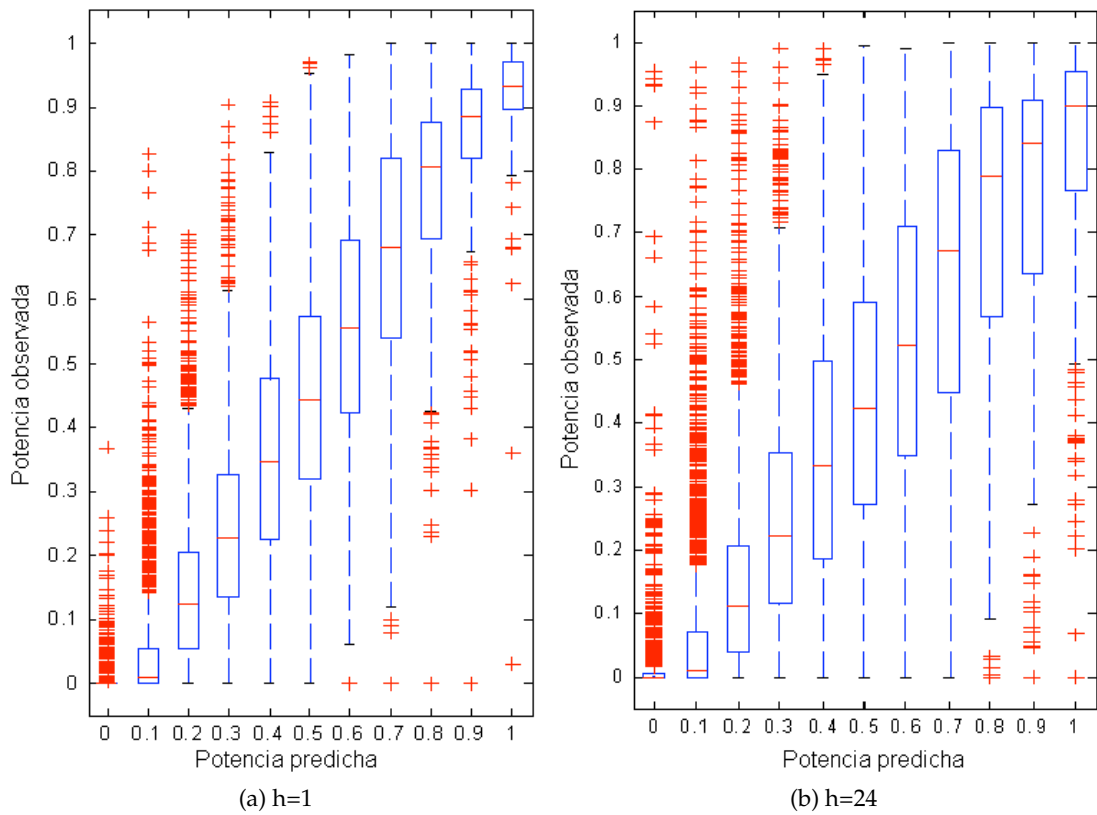


Figura 4.2: Gráfico de caja para de la potencia medida respecto a diferentes niveles de la predicción puntual

#### 4.1.2 Notación

Para facilitar la lectura del capítulo, vamos a introducir la notación empleada a lo largo del mismo. Dado el carácter estocástico del viento, la potencia eólica generada durante un periodo de tiempo puede interpretarse como un proceso estocástico, cuyos valores observados serán denotados por  $p_t$ . Por simplicidad en la notación siempre usaremos  $p_t$ , tanto para referirnos al proceso como a los valores observados. El objetivo del tra-

bajo será caracterizar a la variable aleatoria  $p_{t+h}$  condicionada a la información en el instante de tiempo  $t$ . Para caracterizar dicha variable aleatoria  $p_{t+h}$  tenemos que conocer su función de distribución  $F_{t+h}$  o su función de densidad  $f_{t+h}$ . Denotaremos a la predicción de la variable aleatoria  $p_{t+h}$  a horizonte  $h$  como  $\hat{p}_{t+h|t}$ . El error de predicción a horizonte  $h$  será  $e_{t+h|t} = p_{t+h} - \hat{p}_{t+h|t}$ , y su momento condicionado de orden 1 será  $m_{t+h|t}^{(0)} = E_t [e_{t+h|t}]$ . Los momentos centrales del error de predicción  $e_{t+h|t}$  serán  $m_{t+h|t}^{(j)} = E_t \left[ \left( e_{t+h|t} - m_{t+h|t}^{(0)} \right)^j \right]$ , donde  $j = 2, 3, \dots$

## 4.2 Densidad predictiva en energía eólica

La estimación de la densidad predictiva de la potencia eólica es un área de investigación que se ha empezado a estudiar en años recientes. Existen varios puntos de vista desde los que abordar el problema.

### Empleando la distribución de la velocidad del viento

El principal componente de la potencia eólica producida por un parque es la velocidad del viento, por lo tanto parece evidente que el estudio de la velocidad del viento será una herramienta adecuada para la modelización de la potencia generada a través de predicciones de la velocidad. En el trabajo de Carta et al. (2009) se hace una gran revisión sobre diferentes estimaciones de distribuciones para la velocidad del viento.

Entonces, una vez estimada la distribución de la velocidad del viento, la predicción probabilística de la potencia eólica se construye a partir de la curva de potencia que mide la relación existente entre velocidad y potencia (p.ej., Lange 2005). El principal problema es que dicha relación es no lineal y cambiará dependiendo del tiempo, del aerogenerador, de la temperatura ambiental, de la dirección del viento, etc.

### Empleando "ensemble predictions"

En meteorología es habitual suministrar la predicción de diferentes variables ambientales para un periodo de tiempo en forma de un conjunto de trayectorias posibles a partir de diferentes condiciones iniciales y empleando diferentes modelos (Epstein, 1969). A ese conjunto de trayectorias se le conoce como "ensemble predictions". Entonces, una vez generadas las predicciones *ensemble* la densidad predictiva de la producción de energía eólica se calcula empleando las curvas de potencia que relacionan las variables meteorológicas empleadas con la potencia eólica (Nielsen et al. 2004, Pinson y Madsen 2009, Taylor et al. 2009). En este caso, además del problema de la relación no lineal existente

comentado en el anterior punto, se le añade la dificultad de obtener las predicciones *ensemble*, ya que tienen que ser suministradas por una agencia meteorológica.

### **Empleando regresión cuantílica**

Estos métodos consisten en aplicar una regresión por cuantiles sobre la producción eólica en función de un conjunto de variables explicativas sin tener que asumir ningún tipo de estructura en la densidad condicionada de la potencia. Bremnes (2006) comprobó su eficiencia comparándolo con un modelo Gaussiano y un estimador de Nadaraya-Watson para la función de distribución, obteniendo resultados muy parejos para todos los métodos, dando ventaja al método de Nadaraya-Watson por cuestiones de sencillez en la computación. En Moller et al. (2008) se propone un procedimiento adaptativo para la regresión por cuantiles que mejora el tiempo de computación.

### **Empleando la predicción de potencia eólica**

Otra de las opciones existentes consiste en ajustar una distribución de probabilidad condicionada de la potencia a partir de los , históricos de la potencia medida y de predicciones de la misma. La principal dificultad está en que la potencia eólica es una variable con soporte acotado y no gaussiana, además de ser una variable con un comportamiento no estacionario. Por ello, Lau y McSharry (2010) y Pinson (2010) proponen realizar una transformación logística, con el propósito de conseguir una distribución Normal, empleando a continuación diferentes modelos paramétricos para realizar las predicciones puntuales y estimar la densidad predictiva empleando las funciones de distribución Normal Truncada y Normal Censurada.

Otra opción es ajustar una distribución paramétrica condicionada a partir de la estimación de momentos condicionados de la potencia. Bludszuweit et al. (2008) y Al-Awami y El-Sharkawi (2009) lo hicieron empleando distribuciones como la Beta y la distribución generalizada del valor extremo, a partir de los dos primeros momentos.

En este trabajo seguiremos este enfoque. Para ello proponemos emplear un método adaptativo que nos permite estimar los momentos centrales condicionados de los errores de predicción mediante modelos de regresión adaptativa. Las variables explicativas de estas regresiones serán las predicciones puntuales de la potencia. Una vez obtenidos los momentos, los emplearemos para estimar la densidad predictiva de dichas predicciones empleando para ello diferentes distribuciones paramétricas, como son la Normal Truncada, la Normal Censurada, la Beta y la distribución de Máxima Entropía (Jaynes, 1957). Las diferencias con trabajos anteriores es el uso de la distribución de Máxima Entropía que nos permitirá emplear momentos de hasta orden 4 y que la estimación de

dichos momentos es realizada de forma recursiva y con parámetros que cambian con el tiempo.

### 4.3 Estimación de los momentos condicionados

Para realizar una estimación paramétrica de la densidad predictiva será necesario realizar la estimación de los momentos condicionados. La estimación de los momentos condicionados será efectuada mediante modelos de regresión en los cuales los regresores serán las predicciones puntuales de la potencia eólica.

Entonces, el modelo de regresión empleado para estimar los momentos condicionados es

$$e_{t+h|t}^M = \alpha_{0,t+h} + \alpha_{1,t+h}\hat{p}_{t+h|t} + \alpha_{2,t+h}\hat{p}_{t+h|t}^2 + \alpha_{3,t+h}\hat{p}_{t+h|t}^3 + v_{t+h}, \quad (4.1)$$

donde  $M = 1, 2, \dots$ . De esta manera para diferentes valores de  $M$  se ajustan modelos para diferentes momentos, puesto que el segundo momento está asociado al error de predicción al cuadrado y así sucesivamente.

Suponiendo que las predicciones de energía eólica han sido estimadas empleando un modelo correcto, se espera que la media condicionada del error de predicción sea cero, es decir,  $m_{t+h|t}^{(0)} = E[e_{t+h|t}] = 0$ . En este caso, los momentos centrales condicionados serán

$$m_{t+h}^{(M)} = E_t \left[ e_{t+h|t}^M \right], \quad (4.2)$$

es decir, son calculados directamente empleando el modelo de regresión descrito en la expresión (4.1) para diferentes valores de  $M$ . No obstante, en la práctica será fácil que un predictor de la potencia eólica sea condicionalmente sesgado durante ciertos periodos de tiempo, es decir,  $m_{t+h|t}^{(0)} \neq 0$ . Por ello, se realiza una corrección del sesgo, por lo que para calcular los momentos centrales se emplea el modelo

$$\tilde{e}_{t+h|t}^M = \alpha_{0,t+h} + \alpha_{1,t+h}\hat{p}_{t+h|t} + \alpha_{2,t+h}\hat{p}_{t+h|t}^2 + \alpha_{3,t+h}\hat{p}_{t+h|t}^3 + v_{t+h}, \quad (4.3)$$

donde  $\tilde{e}_{t+h|t}^M = \left( e_{t+h|t} - m_{t+h|t}^{(0)} \right)^M$  para  $M = 2, 3, \dots$

Para comprobar la relación existente entre los momentos centrales de la potencia eólica generada por un parque respecto a sus predicciones se calcula la media de la estimación de los momentos para diferentes horizontes de predicción  $h = 1, \dots, 48$ . En la figura 4.3 se ve que los intervalos de confianza Bootstrap de la estimación nos confirman que los momentos siguen un patrón claro en función del valor previsto, por lo que, aparentemente, las predicciones de potencia serán buenos regresores.



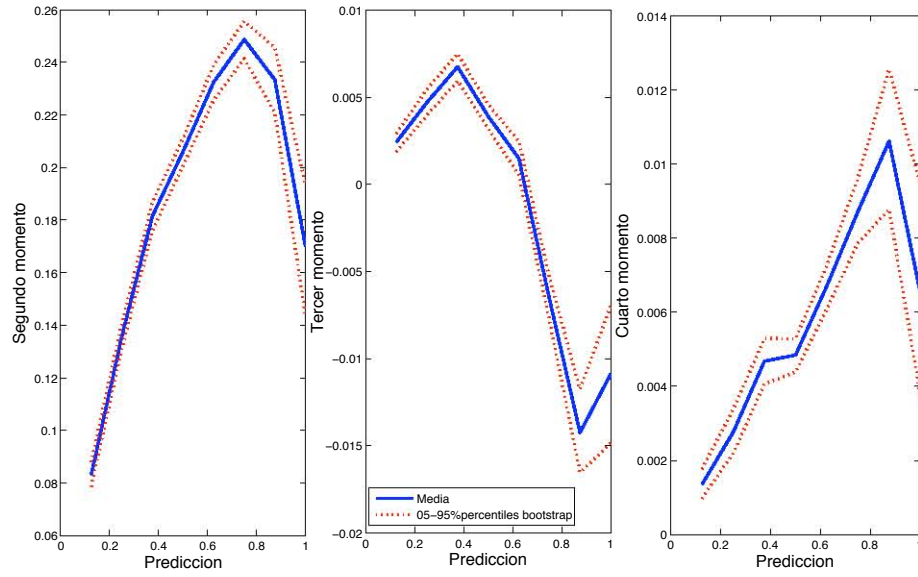


Figura 4.3: Media de la estimación de los momentos para diferentes horizontes

#### 4.3.1 Método de estimación recursiva

El objetivo del método propuesto es ir calculando la senda de distribuciones predictivas  $f_{t+h}$  para un horizonte  $h$ , es decir, habrá que ir estimando de manera recursiva los momentos condicionados. Para ello se ha desarrollado un procedimiento adaptativo.

La estimación del modelo (4.3) se realiza de manera adaptativa empleando el método de mínimos cuadrados recursivos. Por simplificar la notación reescribimos (4.3) en forma matricial, tal que

$$y_{t+h} = X'_{t+h} \alpha_{t+h} + v_{t+h}, \quad (4.4)$$

donde el vector  $y_{t+h} = \tilde{e}_{t+h|t}^M$ , la matriz  $X_{t+h} = (1, \hat{p}_{t+h|t}, \hat{p}_{t+h|t}^2, \hat{p}_{t+h|t}^3)'$  contiene las variables explicativas y el vector  $\alpha_{t+h} = (\alpha_{0,t+h}, \alpha_{1,t+h}, \alpha_{2,t+h}, \alpha_{3,t+h})'$  contiene los parámetros a estimar. Entonces, empleando el método de mínimos cuadrados ponderados de manera recursiva la estimación de la expresión (4.4) será

$$\hat{\alpha}_{t+h} = \hat{\alpha}_{t+h-1} + M_{t+h}^{-1} X_{t+h} \hat{v}_{t+h}, \quad (4.5)$$

donde  $\hat{v}_{t+h} = y_{t+h|t} - X'_{t+h} \hat{\alpha}_{t+h-1}$  es el error de predicción un paso adelante y la matriz de ganancia  $M_{t+h}^{-1}$  se calcula de manera recursiva según

$$M_{t+h}^{-1} = \frac{1}{\lambda_{t+h}} \left( M_{t+h-1}^{-1} - \frac{M_{t+h-1}^{-1} X_{t+h} X'_{t+h} M_{t+h-1}^{-1}}{\lambda_{t+h} + X'_{t+h} M_{t+h-1}^{-1} X_{t+h}} \right), \quad (4.6)$$

donde  $\lambda_{t+h}$  es el factor de olvido que permitirá que la estimación recursiva vaya cambiando a lo largo del tiempo. El factor de olvido empleado en este trabajo es el propuesto por Sánchez (2006a), que está basado en la distancia de Cook.

Como la variable aleatoria con la que trabajamos es la potencia eólica hemos de acotar su estimación para no obtener valores fuera del intervalo  $[0, 1]$ . Es fácil comprobar que las acotaciones serán

$$\hat{e}_{t+h|t}^M \in \begin{cases} \left[ -\hat{p}_{t+h|t}^M, (1 - \hat{p}_{t+h|t})^M \right] & , \text{ si } M \text{ es impar} \\ \left[ 0, (1 - \hat{p}_{t+h|t})^M \right] & , \text{ si } M \text{ es par} \end{cases} . \quad (4.7)$$

En la figura 4.4 podemos ver la senda de estimaciones de los 4 primeros momentos centrales durante un periodo de tiempo, para diferentes horizontes de predicción. Además en la figura 4.5 podemos ver la asimetría y curtosis asociadas a dichos momentos centrales. En alguno de los gráficos hemos tenido que emplear diferentes escalas en los ejes de las ordenadas, para facilitar ver la evolución de los momentos, puesto que según aumenta el horizonte de predicción, los valores serán mayores.

## 4.4 Modelización adaptativa de densidades predictivas

En la sección 4.1.1 mostramos que la distribución de la potencia tiene varias características que deberemos de tener en cuenta a la hora de seleccionar las distribuciones con las que trabajaremos, como son que es acotada, que es unimodal, que es condicionalmente asimétrica, etc. Como la potencia producida está acotada no se puede emplear la distribución Normal, por ello hemos empleado dos posibles opciones basadas en la Normal como son la Normal Truncada y la Normal Censurada. Estas distribuciones normales con soporte acotado presentan ya diferentes asimetrías y curtosis que podrían ser adecuadas para nuestro problema.

Otra distribución apropiada para nuestros , será la Beta, ya que está acotada en  $[0, 1]$  y se puede ajustar a fenómenos asimétricos. Por último, emplearemos la distribución de Máxima Entropía, ya que también nos permitirá obtener una densidad acotada y nos es muy flexible para adaptarse a momentos condicionados cambiantes.

### 4.4.1 Distribución Normal truncada

Sea  $\phi(\cdot)$  la función de densidad de una distribución Normal de media  $\mu$  y desviación  $\sigma$  y  $\Phi(\cdot)$  su correspondiente función de distribución. Entonces, la variable  $y \in R = [a, b]$ , donde  $-\infty < a < b < \infty$ , sigue una distribución Normal Truncada de densidad

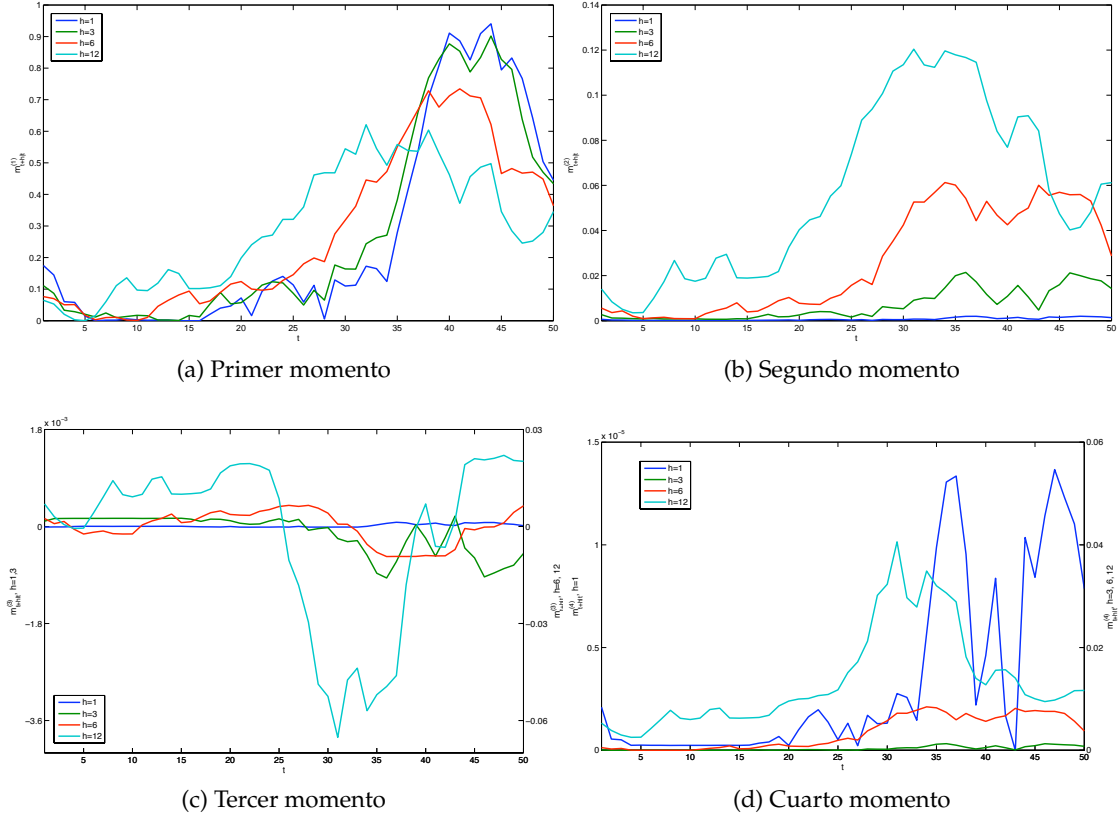


Figura 4.4: Estimación recursiva de los momentos centrados condicionados para  $h=1,3,6,12$ .

$$f(y) = \frac{\frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \quad a < y < b. \quad (4.8)$$

Entonces, los momentos de la variable truncada podrán ser calculados fácilmente a partir de los momentos de la variable sin truncar  $\mu$  y  $\sigma$  (p.ej. Cohen 1961),

$$\mu_{(NT)} = \mu + \sigma\delta; \quad (4.9)$$

$$\sigma_{(NT)}^2 = \sigma^2 \left[ 1 - \delta^2 - \frac{[a - \mu] \phi\left(\frac{a-\mu}{\sigma}\right) - [b - \mu] \phi\left(\frac{b-\mu}{\sigma}\right)}{\sigma \left[ \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right]} \right], \quad (4.10)$$

donde

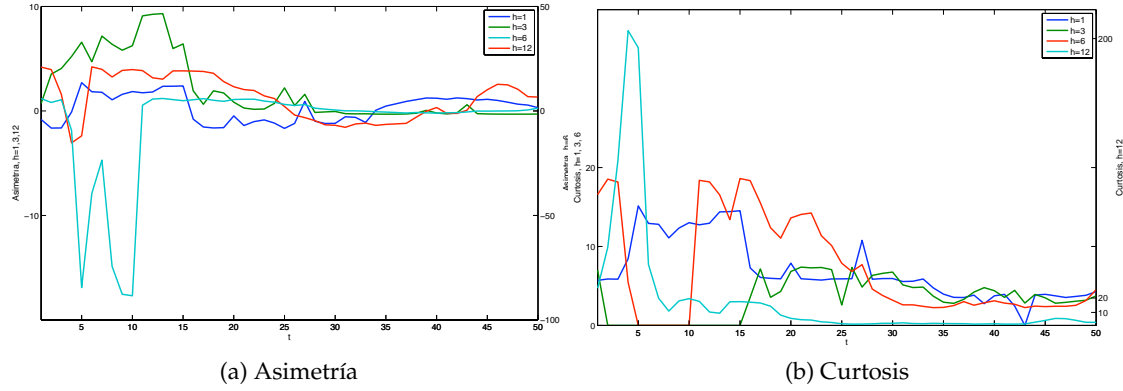


Figura 4.5: Estimación recursiva de los momentos condicionados para  $h=1,3,6,12$ .

$$\delta = \frac{\phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}. \quad (4.11)$$

### Estimación de la densidad predictiva

Empleando la expresión (4.8) se puede estimar la función de densidad Normal Truncada empleando para ello los momentos de la variable sin truncar. El problema es que calcular los momentos de la variable truncada a partir de los de la variable sin truncar, como vimos en (4.9) y (4.10), es directo, pero la dirección contraria, que justamente es la que necesitamos, no lo es.

Si asumimos que la distribución de las predicciones empleadas en nuestra estimación es una Normal Truncada, estimamos los dos primeros momentos condicionados empleando el modelo (4.3), con  $M = 1, 2$ . El problema es que de esta manera obtenemos los momentos de la variable truncada  $\mu_{(NT)}$  y  $\sigma_{(NT)}$  y no los momentos de la variable sin truncar que necesitaremos para estimar la densidad  $f_{t+h}^{(NT)}$ .

Cohen (1991) muestra que los momentos sin truncar pueden ser estimados mediante

$$\hat{\sigma} = \frac{b - a}{\hat{\xi}_2 - \hat{\xi}_1}; \quad (4.12)$$

$$\hat{\mu} = a - \hat{\sigma}\hat{\xi}_2, \quad (4.13)$$

donde

$$\xi_1 = \frac{a - \mu}{\sigma} \text{ y } \xi_2 = \frac{b - \mu}{\sigma}. \quad (4.14)$$

La estimación de  $\hat{\xi}_1$  y  $\hat{\xi}_2$  puede ser efectuada mediante diferentes métodos iterativos

como el de Newton-Raphson, pero Cohen (1991) propone emplear un procedimiento gráfico desarrollado en Cohen (1957).

Entonces, empleando las expresiones (4.9) y (4.10) se obtienen los momentos sin truncar  $\mu$  y  $\sigma$  resolviendo el sistema mediante un algoritmo de búsqueda. De esta manera, empleamos la expresión (4.8) para obtener la estimación de la distribución predictiva  $f_{t+h}^{(NT)}$ . En la figura (4.6) podemos ver un ejemplo de la senda de estimaciones  $f_{t+1}^{(NT)}$  para diferentes periodos de tiempo.

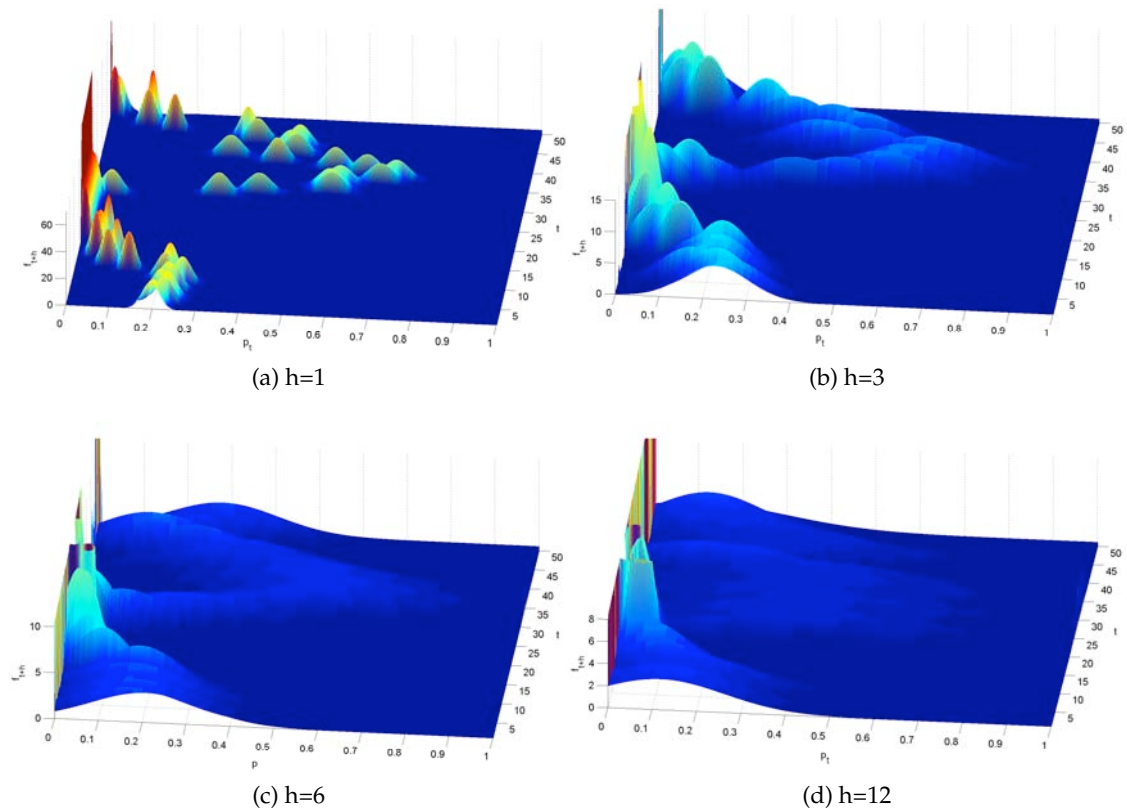


Figura 4.6: Ejemplo de la senda de densidades predictiva empleando la distribución Normal Truncada para diferentes horizontes

#### 4.4.2 Distribución Normal censurada

Sea  $y$  una variable aleatoria con función de densidad  $\phi(\cdot)$  Normal de media  $\mu$  y desviación  $\sigma$  y  $\Phi(\cdot)$  su correspondiente función de distribución. Entonces, la variable  $y' \in R = [a, b]$ ,

donde  $-\infty < a < b < \infty$ , sigue una distribución Normal Censurada donde

$$y' = \begin{cases} a & , \text{ si } y \leq a \\ N(\mu, \sigma) & , \text{ si } a < y < b \\ b & , \text{ si } y \geq b \end{cases} \quad (4.15)$$

En este caso, las colas de la Normal que sobrepasen el intervalo en el que la variable está acotada se acumulan en los extremos. Los momentos de la variable censurada son (Cohen, 1991)

$$\mu_{(NC)} = a\Phi(\xi_1) + [\Phi(\xi_2) - \Phi(\xi_1)](\mu + \sigma\delta) + [1 - \Phi(\xi_2)]b. \quad (4.16)$$

$$\begin{aligned} \sigma_{(NC)}^2 = & a^2\Phi(\xi_1) + [\Phi(\xi_2) - \Phi(\xi_1)](\sigma^2 + \mu^2 + 2\mu\sigma\delta) + \\ & \sigma[[a - \mu]\Phi(\xi_1) - [b - \mu]\Phi(\xi_2)] + [1 - \Phi(\xi_2)]b^2 - \mu_{(NC)}^2, \end{aligned} \quad (4.17)$$

donde

$$\delta = \frac{\phi(\xi_2) - \phi(\xi_1)}{\Phi(\xi_2) - \Phi(\xi_1)}. \quad (4.18)$$

### Estimación de la densidad predictiva

En este caso, nos encontramos en el mismo problema que para la estimación de la distribución truncada. Además, Cohen (1991) muestra cómo la estimación de los momentos de la variable sin censurar se hará mediante las mismas expresiones que en la distribución Normal Truncada, es decir,

$$\hat{\sigma} = \frac{b - a}{\hat{\xi}_2 - \hat{\xi}_1}; \quad (4.19)$$

$$\hat{\mu} = a - \hat{\sigma}\hat{\xi}_2, \quad (4.20)$$

solo que la estimación de  $\hat{\xi}_1$  y  $\hat{\xi}_2$ , en esta ocasión será diferente. Para hallar los momentos de la variable sin censurar, dicha estimación tendrá que ser realizada mediante un procedimiento iterativo descrito en Cohen (1957, 1991).

El primer paso importante en un proceso iterativo es obtener una buena estimación inicial que esté próxima a la solución óptima, para que de esta manera el número de iteraciones sea el menor posible. Como valores iniciales  $\hat{\xi}_1^0$  y  $\hat{\xi}_2^0$ , Cohen (1957) emplea

$$\Phi(\hat{\xi}_1^0) = \frac{c_1}{N} \text{ y } 1 - \Phi(\hat{\xi}_2^0) = \frac{c_2}{N}, \quad (4.21)$$

donde  $c_1$  y  $c_2$  son las el número de observaciones muestrales censuradas por la izquierda y por la derecha respectivamente, y  $N$  es el tamaño total de la muestra. Una vez hallados los valores iniciales Cohen (1957) recomienda mejorar la estimación realizando un par de iteraciones empleando las "sustituciones sucesivas de Cohen", donde

$$\begin{aligned}\hat{\xi}_1^{i+1} &= \frac{\hat{\xi}_2^{i+1} v_1/w - (\Omega_1^i - \Omega_2^i)}{1 - v_1/w} \\ \hat{\xi}_2^{i+1} &= (\Omega_1^i - \Omega_2^i) + \frac{\left\{ [(\Omega_1^i - \Omega_2^i) v_1/w + \Omega_2^i] - \sqrt{[(\Omega_1^i - \Omega_2^i) v_1/w + \Omega_2^i]^2 + 4s^2/w^2} \right\} w(v_1 - w)}{2s^2}\end{aligned}\quad (4.22)$$

donde  $v_1 = \mu_{(NC)} - a$ ,  $w = b - a$ ,  $s^2 = \sigma_{(NC)}^2$  y

$$\begin{aligned}\Omega_1^i &= \frac{c_1}{n} \frac{\phi(\hat{\xi}_1^i)}{\Phi(\hat{\xi}_1^i)} \\ \Omega_2^i &= \frac{c_2}{n} \frac{\phi(\hat{\xi}_2^i)}{1 - \Phi(\hat{\xi}_2^i)}\end{aligned}\quad (4.23)$$

siendo  $n$  el número de observaciones muestrales sin censurar. A partir de entonces, se emplea un método de Newton hasta alcanzar el óptimo. Para ello, se usan los valores  $\hat{\xi}_1^{i+1}$  y  $\hat{\xi}_2^{i+1}$  obtenidos para estimar los valores iniciales del método de Newton  $\hat{\mu}^0$  y  $\hat{\sigma}^0$ , y a partir de ellos se buscarán las estimaciones de los momentos muestrales sin censurar.

Es importante notar que al estimar los momentos empleando los datos que poseemos, obtendremos los momentos correspondientes o bien a la muestra truncada o bien a la censurada, dependiendo del caso en el que estemos. Por ello, es necesario resolver el sistema que nos permita calcular el valor de los momentos  $\mu$  y  $\sigma$  de la distribución sin truncar o sin censurar. No es infrecuente encontrar trabajos donde se ignora este problema.

En la figura 4.7 podemos ver un ejemplo de la senda de estimaciones  $f_{t+1}^{(NC)}$  para diferentes periodos de tiempo.

#### 4.4.3 Distribución Beta

Sea  $y$  una variable que sigue una distribución  $Beta(\alpha, \beta)$ , entonces su función de densidad es

$$f(y) = \frac{y^{\alpha-1} [1-y]^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < y < 1, \quad (4.24)$$

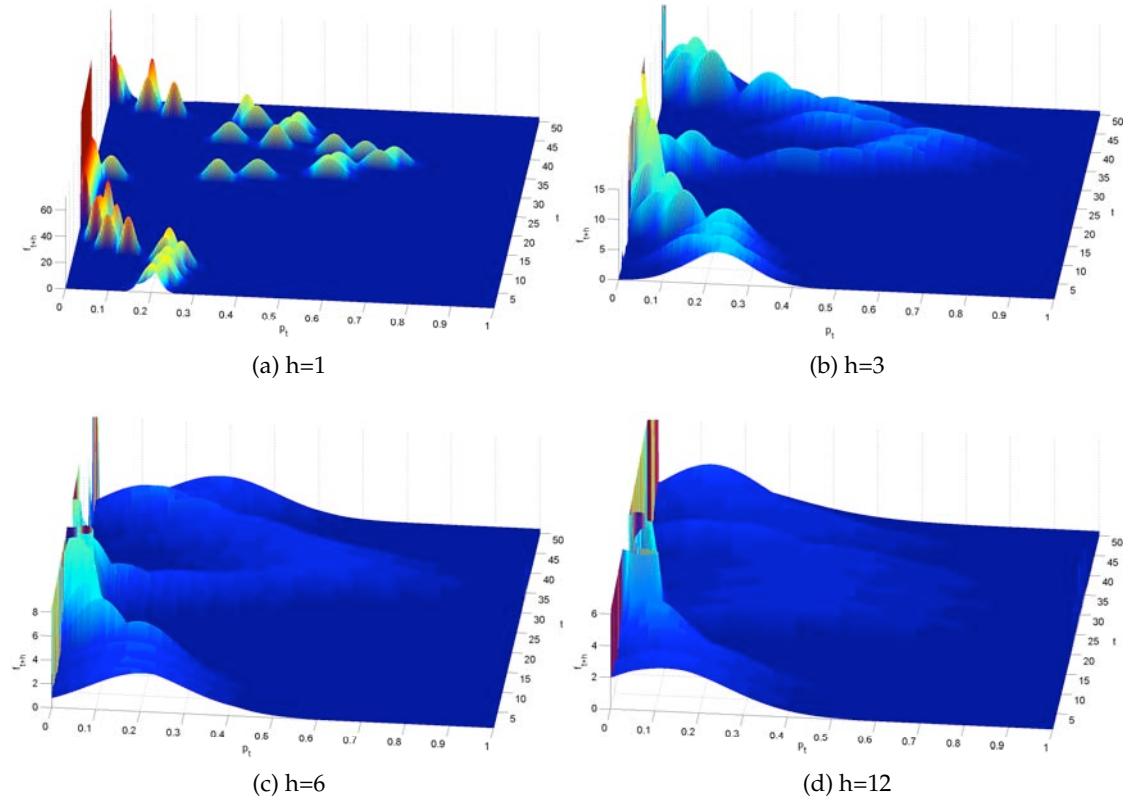


Figura 4.7: Ejemplo de la senda de densidades predictiva empleando la distribución Normal Censurada para diferentes horizontes

donde  $B(\alpha, \beta)$  es la función Beta. Los parámetros  $\alpha$  y  $\beta$  se estimarán, por ejemplo, empleando los dos primeros momentos de la variable  $y$  a partir de la relación

$$\alpha = \frac{\mu^2 - \mu\sigma^2}{\sigma^2 - \mu^2} ; \quad \beta = \frac{\mu - \sigma^2}{\sigma^2 - \mu^2} - \alpha \quad (4.25)$$

### Estimación de la densidad predictiva

En este caso la estimación será sencilla, ya que empleando la expresión (4.3), con  $M = 1, 2$ , obtendremos la estimación de los dos primeros momentos, que nos permiten calcular los parámetros de la distribución Beta mediante la expresión (4.25). Entonces, la función de densidad Beta  $f_{t+h}^{(B)}$  se calcula empleando (4.24). En la figura 4.8 podemos ver un ejemplo de la senda de estimaciones  $f_{t+1}^{(B)}$  para diferentes periodos de tiempo.



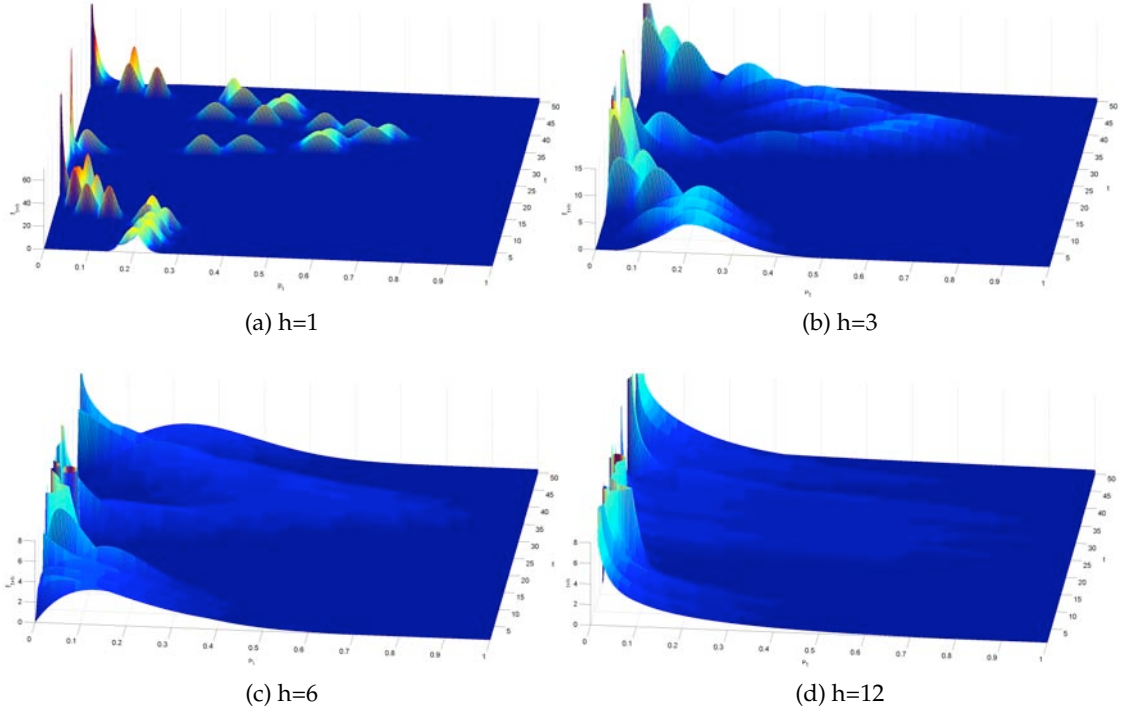


Figura 4.8: Ejemplo de la senda de densidades predictiva empleando la distribución Beta para diferentes horizontes

#### 4.4.4 Distribución de Máxima Entropía

La entropía puede ser vista como una medida de la incertidumbre para una función de densidad (Shannon ,1948)

$$S = - \int_R f(y) \ln(f(y)) dy. \quad (4.26)$$

Jaynes (1957) empleó este concepto para proponer una distribución de probabilidad que minimice la información previa necesaria, que será aquella que maximice la entropía de Shannon. Entonces, la función de densidad de máxima entropía  $f(y)$  definida en el intervalo  $[a, b]$  es aquella que cumpla

$$\begin{aligned} \max \quad & - \int_a^b f(y) \ln(f(y)) dy \\ \text{s.a.} \quad & \int_a^b f(y) dy = 1 \\ & \int_a^b y^i f(y) dy = m_i, \quad i = 1, \dots, M, \end{aligned} \quad (4.27)$$

donde los  $m_i$  son los  $M$  primeros momentos respecto al origen. Jaynes (1957) muestra que la forma analítica de la función de densidad de máxima entropía  $f(y)$  es

$$f(y) = \exp \left[ \lambda_0 + \sum_{i=1}^M \lambda_i y^i \right], \quad (4.28)$$

donde

$$\lambda_0 = -\ln \left[ \int_a^b \exp \left( \sum_{i=1}^M \lambda_i y^i \right) dy \right], \quad (4.29)$$

y los multiplicadores  $\lambda_1, \dots, \lambda_M$  son calculados resolviendo el sistema de  $M$  ecuaciones no lineales

$$\left[ \int_a^b y^j \exp \left( \sum_{i=1}^M \lambda_i y^i \right) dy \right] \left[ \int_a^b \exp \left( \sum_{i=1}^M \lambda_i y^i \right) dy \right]^{-1} = m_j, \quad j = 1, \dots, M. \quad (4.30)$$

El sistema descrito en (4.30) solo tiene solución analítica cuando  $M = 1$ . En este trabajo llegaremos a emplear un  $M = 4$ , por lo que el sistema (4.30) tendrá que ser resuelto numéricamente. El procedimiento que hemos seguido para ello es el descrito por Siddall y Diab (1975). Los pasos del procedimiento son:

- (i) Calcular los  $M = 4$  primeros momentos condicionados centrales de manera adaptativa.
- (ii) Transformar los  $M$  momentos centrales  $m_j$  en momentos en el origen  $m_j^{(0)}$ , empleando
$$\begin{aligned} m_1^{(0)} &= m_1 \\ m_j^{(0)} &= m_j - m_1^j (-1)^j - \sum_{k=1}^{j-1} \frac{(-1)^k j!}{k!(j-k)! m_1^k m_{j-k}^{(0)}}. \end{aligned} \quad (4.31)$$
- (iii) Elegir los valores iniciales  $\lambda_1^0, \dots, \lambda_M^0$ . Siddall y Diab (1975) proponen diferentes alternativas, entre las que hemos escogido la más conservadora  $\lambda_1^0 = \dots = \lambda_M^0 = 0$ .
- (iv) Calcular los multiplicadores  $\lambda_1, \dots, \lambda_M$  resolviendo el problema de programación no lineal sin restricciones

$$1 - \frac{\sum_{i=1}^l a_i x_i^j \exp \left( \sum_{j=1}^M \lambda_j x_i^j \right)}{m_j^0 \sum_{i=1}^l a_i \exp \left( \sum_{j=1}^M \lambda_j x_i^j \right)} = R_j, \quad j = 1, \dots, M, \quad (4.32)$$

donde  $a_i$  son los multiplicadores de integración numérica,  $l$  es el número de intervalos empleados en dicha integración numérica y  $R_j$  son los residuos normalizados empleados en la resoluciones de las ecuaciones simultáneas (4.32).

- (v) Emplear la expresión (4.29) para calcular  $\lambda_0$ .
- (vi) Emplear la expresión (4.28) para estimar la distribución de máxima entropía.

El número de momentos que hemos empleado es  $M = 4$ , pero en ocasiones el problema de programación no lineal (4.32) tendrá problemas numéricos y no convergerá, por lo que será necesario reducir el número de momentos hasta que converja. La experiencia nos ha mostrado que esto pasa especialmente cuando el segundo momento toma valores aproximadamente iguales a cero. Aun así, el número mínimo de momentos que ha sido empleado es  $M = 2$ . Este problema de convergencia deja claro que un método para seleccionar el número de momentos condicionados a emplear mejoraría la eficacia de la estimación de la distribución de Máxima Entropía.

### Estimación de la densidad predictiva

El método de máxima entropía ha sido utilizado para la estimación de la densidad de la velocidad del viento en trabajos previos (Ramírez y Carta, 2006; Akpınar y Kavak Akpınar, 2007; Liu y Chang, 2010). Sin embargo no ha sido empleado para la modelización de la densidad predictiva de la potencia eólica, a pesar de que su uso tiene varias ventajas para dicho problema. La principal, es que basándonos sólo en los datos y sin necesidad de ningún tipo de asunción previa, podemos estimar una distribución acotada en el dominio  $[a, b]$  que queramos.

Para estimar la densidad predictiva  $f_{t+h}^{(MEP)}$  estimaremos los cuatro primeros momentos condicionados empleando la expresión (4.3), con  $M = 1, 2, 3, 4$ . Entonces, se resuelve de manera numérica el sistema descrito en (4.30), y obtenemos la estimación de la densidad predictiva  $f_{t+h}^{(MEP)}$  mediante (4.28) y (4.29). En la figura 4.9 podemos ver un ejemplo de la senda de estimaciones  $f_{t+1}^{(MEP)}$  para diferentes periodos de tiempo.

### 4.4.5 Comparación de las distribuciones empleadas

En la figura 4.10 podemos ver varios ejemplos de las diferentes estimaciones obtenidas en un determinado instante de tiempo para diferentes horizontes de predicción. Para  $h = 1$  vemos un ejemplo en el que todas las distribuciones funcionan de manera similar puesto que nos encontramos en mitad del intervalo y la desviación es pequeña. En el ejemplo para un  $h = 3$  vemos las diferencias existentes cuando nos encontramos en un extremo, diferencias que se irán acrecentando según aumenta el horizonte y con ello la desviación como vemos en los gráficos restantes.

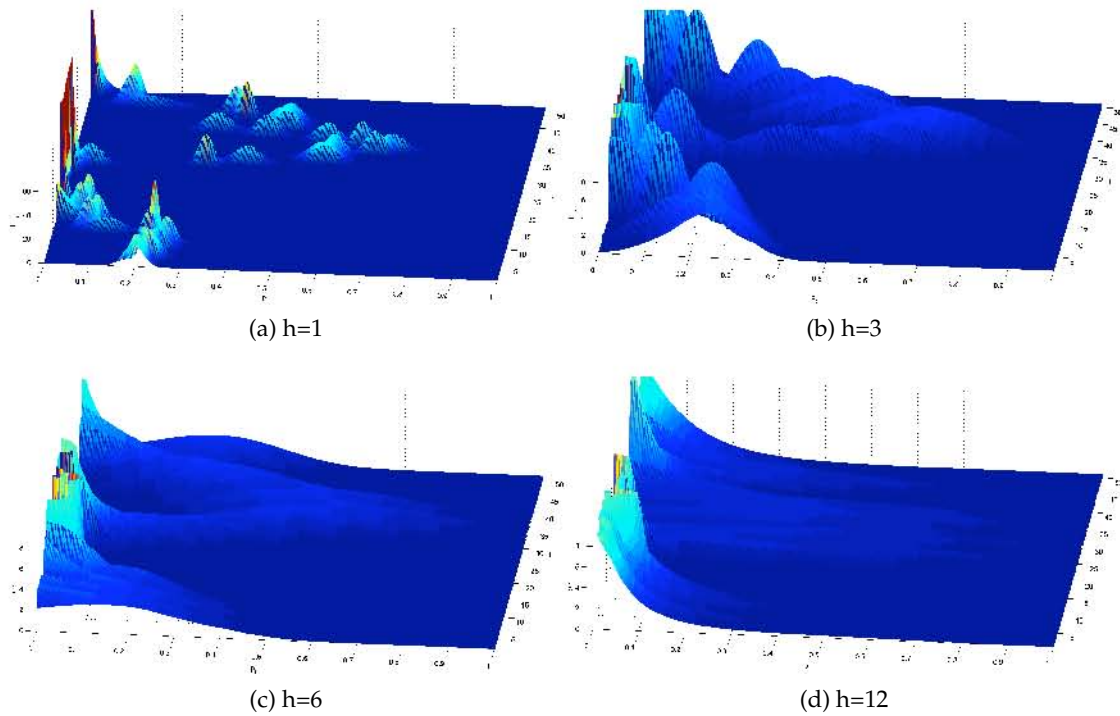


Figura 4.9: Ejemplo de la senda de densidades predictiva empleando la distribución de máxima entropía para diferentes horizontes

## 4.5 Criterios de evaluación de densidades predictivas

Existen multitud de herramientas para evaluar el correcto funcionamiento de una secuencia de estimaciones de densidades predictivas. Se pueden clasificar los diferentes métodos en dos tipos distintos, por una parte están aquellos métodos que sirven para comprobar que la densidad predictiva estimada es correcta. Estos son métodos necesarios, pero no suficientes pues no garantizan un buen comportamiento predictivo futuro. Por ello son necesarios otro tipo de métodos que comparen entre diferentes densidades predictivas y digan cuál es la más eficiente.

### 4.5.1 Comprobación de la estimación

La evaluación de una densidad predictiva se basa en el "*Prequential principle*", llamado así por Dawid (1984). Dawid formó la palabra *Prequential* uniendo los términos secuencia probabilística (*Probabilistic sequential*). La idea del principio es que dado un proceso  $\{Y_{t+h}\}$ , y una secuencia de densidades predictivas  $\{f_{t+h}\}$ , la eficiencia de la estimación de las densidades predictivas podrá ser evaluada empleando únicamente el par  $\{Y_{t+h}, f_{t+h}\}$ .

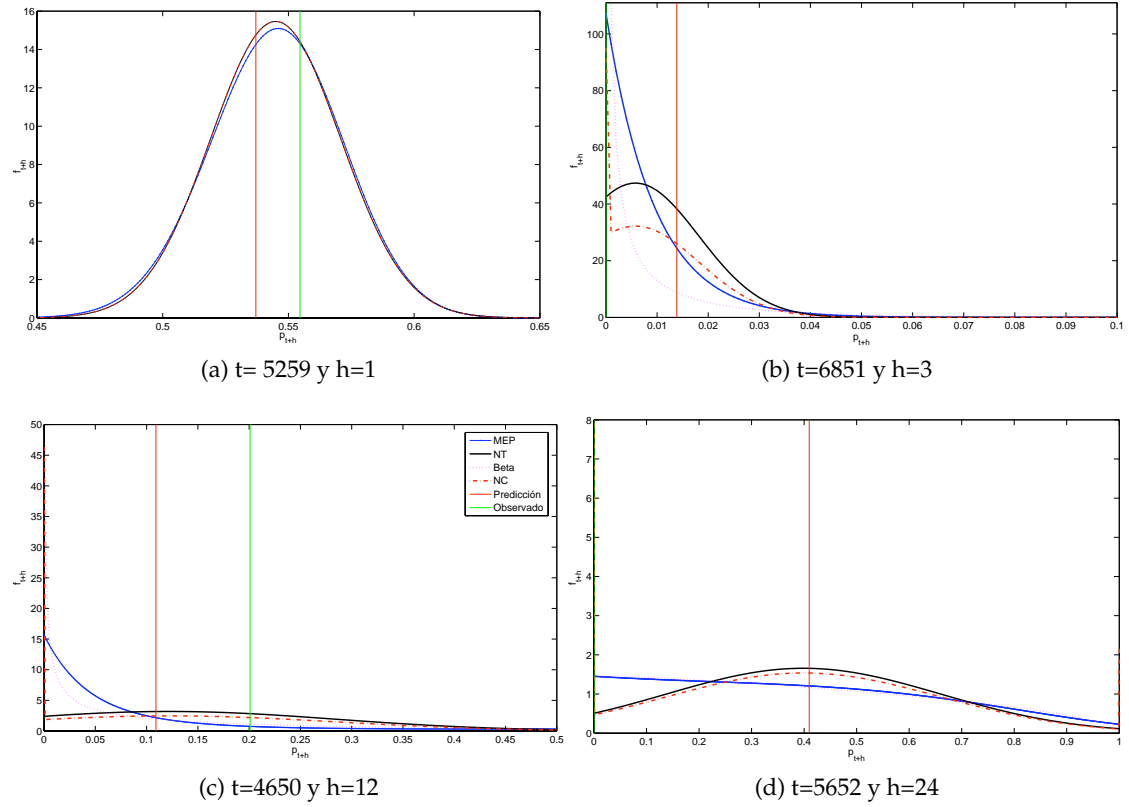


Figura 4.10: Densidades predictivas para diferentes instantes de tiempo y horizontes

Empleando este principio Dawid (1984) propone evaluar si una secuencia de densidades predictivas es adecuada empleando el método PIT (*Probability Integral Transform*). Sean  $u_t$  los valores PIT, definidos como

$$u_{t+h} = \int_{-\infty}^{Y_{t+h}} f_{t+h} dY_t, \quad (4.33)$$

si asumimos que la secuencia de estimaciones  $\{f_{t+h}\}$  es adecuada entonces, Rosenblatt (1952) probó que los valores PIT  $\{u_{t+h}\}$  serán un proceso independiente e idénticamente distribuido según una distribución Uniforme  $U[0, 1]$ . El método PIT propuesto por Dawid (1984) empezó a ser ampliamente usado a partir del trabajo de Diebold et al. (1998) que popularizó su uso. A partir de este trabajo se han propuesto diferentes procedimientos para evaluar diferentes propiedades estadísticas de la secuencia de densidades predictivas se han ido proponiendo. En Corradi y Swanson (2006) se hace una amplia revisión de este tema. Más recientemente González-Rivera et al. (2011) proponen una ampliación para el caso multivariante.

Durante este trabajo, como herramienta para evaluar la secuencia de estimaciones

predictivas hemos empleado el método PIT. Para comprobar que los valores PIT siguen una distribución uniforme hemos realizado, como se hace habitualmente, el histograma de los valores PIT obtenidos. Otra posible representación, quizás más útil para comparar secuencias de densidades predictivas diferentes es obtener un gráfico de los percentiles empíricos del proceso  $\{u_{t+h}\}$  obtenido respecto a los percentiles nominales esperados.

Hamill (2001) muestra un ejemplo del funcionamiento del método PIT como método de evaluación. En el ejemplo, emplea cuatro densidades predictivas diferentes. Una es la ideal y las otras tres son densidades sesgadas. En todos los casos, el método PIT nos dice que las cuatro distribuciones son correctas a pesar de que tres de ellas no lo son. Esto es debido a que el método PIT es una condición necesaria pero no suficiente, por lo que se necesitan emplear otros métodos que permita distinguir cual es el predictor ideal.

#### 4.5.2 Verificación de las predicciones

El concepto de verificación de las predicciones fue introducido en el campo de la meteorología por Jolliffe y Stephenson (2003) y se basa en la idea de obtener medidas que permiten la evaluación de las densidades predictivas. La mayoría de las medidas existentes se basan en el Brier Score, introducido por Brier (1950). Sea  $F_t$  la función de distribución acumulada predictiva de la potencia eólica en el instante de tiempo  $t$  y sea  $p_t$  el valor observado en dicho instante de tiempo, entonces el Brier Score se calcula mediante

$$BS(x) = \frac{1}{N} \sum_{t=1}^N [F_t - H(x - p_t)]^2, \quad (4.34)$$

donde  $H(x - p_t)$  es la función Heaviside y toma valores 0 cuando  $x < p_t$  y valores 1 en caso contrario, siendo  $x$  el dominio de  $p_t$ . El Brier Score es uno de los métodos conocidos como *Scoring Rules*. Estos métodos asignan una puntuación numérica a la densidad predictiva, basandose en el concepto de concentración. Esto es, una distribución predictiva más concentrada, nos dará predicciones mejores. Así pues, un menor valor del Brier Score nos dará una densidad predictiva mejor.

Diferenciar cuál de las densidades predictivas alternativas es mejor puede ser complicado a partir sólo del Brier Score. Por esta razón, es aconsejable una medida que nos diga cual de las distribuciones acumulará menor valor Brier Score. Existen multitud de medidas que permiten esta tarea. En Gneiting y Raftery (2007) podemos ver una amplia revisión. En este trabajo hemos empleado la medida propuesta por Matheson y Winkler (1976), el CRPS (*Continuous Ranked Probability Score*) que vendrá definido por

$$CRPS = \int_{-\infty}^{\infty} BS(x) \, dx, \quad (4.35)$$

la cual, al ser una medida asociada al Brier Score, nos dará como mejor densidad predictiva a aquella que tenga un menor valor CRPS.

## 4.6 Selección de la distribución óptima empleando criterios de evaluación

En la sección anterior, vimos los resultados de las diferentes funciones de distribución que hemos considerado. Sin embargo, es posible que a lo largo del tiempo la que consideramos mejor densidad predictiva vaya cambiando entre las diferentes opciones que consideramos. Por ello, proponemos emplear los criterios de evaluación explicados anteriormente para ir escogiendo la distribución más adecuada dependiendo del instante de tiempo.

Para seleccionar en un determinado instante de tiempo la densidad predictiva más adecuada hemos tenido en cuenta dos posibles criterios:

- (i) Mínimo Brier Score ( $f_{t+h}^{(\min BS)}$ ). En el instante de tiempo  $t + h$  seleccionaremos como densidad predictiva aquella que tenga menor valor del Brier Score predicho en  $\hat{p}_{t+h|t}$ .
- (ii) Mínimo CRPS ( $f_{t+h}^{(\min CRPS)}$ ). En el instante de tiempo  $t + h$  seleccionaremos como densidad predictiva aquella que minimice el valor de CRPS.

El problema es que para calcular tanto el Brier Score como el CRPS necesitamos el valor observado en el instante de tiempo deseado. Entonces, suponiendo que en el instante de tiempo  $t$  hemos calculado una predicción del BS y del CRPS, y dado que en dicho instante conocemos tanto ambos valores, generaremos la siguiente predicción empleando el alisado simple tipo EWMA (Holt, 1957). Es decir,

$$\widehat{BS}_{t+1} = \lambda \widehat{BS}_t + (1 - \lambda) BS_t,$$

donde  $0 < \lambda < 1$  determinará el peso que le damos a la predicción anterior. Un valor muy próximo a 1 nos irá dando predicciones similares. Para el CRPS realizaremos la misma operación. Como valor de  $\lambda$  probamos varios valores muy próximos a 1, empleando finalmente un  $\lambda = 0.9$ .

## 4.7 Comportamiento de la estimación en diferentes parques eólicos

Una vez estimadas las diferentes sendas de densidades predictivas empleando las diferentes distribuciones y los criterios de selección anteriormente detallados, pasamos a evaluar las diferentes estimaciones realizadas. Poseemos los valores históricos de la potencia media generada horariamente por 6 parques eólicos. Además, para cada una de estas potencia tenemos la serie de sus predicciones hasta un horizonte de predicción  $h = 24$ . El tamaño de los ficheros varía, siendo 18937, 3649, 2118, 8658, 5047 y 3220.

Comenzaremos viendo los resultados para el parques eólico con más observaciones disponibles, el Parque 1. En las figuras 4.11, 4.12, 4.13 y 4.14 mostramos los histogramas de los valores PIT obtenidos para las diferentes distribuciones ajustadas para el Parque 1 a horizontes de predicción  $h = 1, 6, 12, 24$ . Como dijimos anteriormente si la distribución estimada es correcta esperamos que los valores PIT sigan una distribución uniforme, por lo que el histograma deberá de tener forma de rectángulo. Como vemos en los diferentes gráficos la distribución con mayores problemas es la Normal Truncada, puesto que acumula un gran número de valores próximos al cero. Las demás distribuciones en general, funcionan de manera correcta, especialmente la distribución MEP y las densidades estimadas mediante los criterios de selección detallados en la sección anterior.

Mediante la representación de los valores PIT podemos ver si las distribuciones estimadas son adecuadas, pero no podemos comprobar su eficacia. Para ello empleamos los métodos para la verificación de densidades predictivas. Una opción es comparar los diferentes valores que toma el Brier Score para las distintas distribuciones empleadas, siendo la mejor distribución aquella que menor Brier Score tenga. En la figura 4.15 vemos los resultados para diferentes horizontes de predicción. Como vemos a horizontes 1 y 6 la distribución MEP obtiene menores valores, pero a horizontes mayores los resultados son más confusos puesto que las diferentes distribuciones obtienen valores similares.

Para comprobar de manera más exacta el resultado observado con la representación del Brier Score, vemos los valores que toma el CRPS. En la figura 4.16a vemos cómo según aumenta el horizonte de predicción la diferencia entre las distintas distribuciones se va reduciendo. En general, podemos concluir que la mejor distribución empleada es la MEP, puesto que tiene el valor mínimo excepto a horizonte 24 en el cual la distribución *Min CRPS* obtiene un valor ligeramente menor.

Para resumir los resultados obtenidos en los 6 parques eólicos empleados vamos a agupar los resultados en un único gráfico por horizonte. Para ello calculamos los



percentiles de los valores PIT estimados para cada distribución para cada uno de los parques y a continuación hallamos la diferencia de la media de los 5 parques restantes respecto a los percentiles nominales. En la figura 4.17 vemos los resultados obtenidos. Como puede verse en estas figuras, según aumenta al horizonte de predicción casi todas las distribuciones van alejándose del cero, es decir la diferencia entre los percentiles empíricos y el óptimo se va agrandando. Sin embargo, la distribución del MEP permanece siempre próximo al cero, confirmando los resultados observados en el Parque 1. Por último, en la figura 4.16 podemos ver los valores del CRPS obtenidos para los diferentes parques eólicos. Como vemos la mejor distribución es la MEP, puesto que, en general, obtiene el menor valor en los diferentes parques. Vemos además resultados prometedores empleando los criterios de selección de distribuciones, especialmente empleando el mínimo CRPS. Como vemos asumir un tipo de distribución a este tipo de datos no parece buena idea, puesto que dependiendo del parque la Normal Truncada, la Normal Censurada o la Beta puede dar buenos o malos resultados, dependiendo de los datos.

## 4.8 Conclusiones y líneas futuras

Calcular la densidad predictiva de las predicciones es siempre un problema complicado, y más aun si la variable a estudiar tiene las características de la potencia eólica. Como vemos en los resultados mostrados asumir que la potencia eólica sigue una distribución conocida no parece buena idea, por lo que el uso de una distribución como la de Máxima Entropía o el uso de un criterio que vaya seleccionando la distribución óptima en cada instante de tiempo parecen ideas que mejoran las propuestas previas. Además, la estimación de los momentos centrales es un problema básico para obtener una buena estimación de la densidad predictiva, y en las propuestas previas dicha estimación se limitaba al uso de la estimación muestral de los momentos.

Los tres puntos claves que proponemos en este tema son, por lo tanto, la estimación adaptativa de los momentos condicionados, el uso como densidad predictiva de la distribución de Máxima Entropía y el uso de criterios de selección para emplear la distribución más adecuada en un instante de tiempo determinado.

En cuanto a la estimación adaptativa, se han realizado ya diferentes pruebas con modelos semejantes empleando diferentes potencias y los resultados son similares. Una posible ampliación sería añadir como regresores a los modelos predicciones pasadas y no solo las predicciones presentes como empleamos actualmente.

La estimación de la distribución de Máxima Entropía conlleva resolver un problema de programación no lineal en cada instante de tiempo, por ello una gran mejora computacional sería obtener una forma recursiva de dicho algoritmo, puesto que al ser una

distribución condicionada esta irá cambiando dependiendo del valor anterior. Otro resultado interesante, sería la obtención de un criterio formal para que el número de momentos empleados en cada instante de tiempo fuera variando según las características de los datos.

Por último, la selección de distribuciones es, como parecen indicar los resultados, un método prometedor. El uso del mínimo CRPS como criterio de selección de distribuciones funciona de manera eficiente sin haber empleado un método de predicción especialmente refinado. Otra idea a explorar será la de realizar una combinación de diferentes distribuciones empleando el Brier Score, que nos dice que distribución es mejor en diferentes tramos del dominio de la variable.

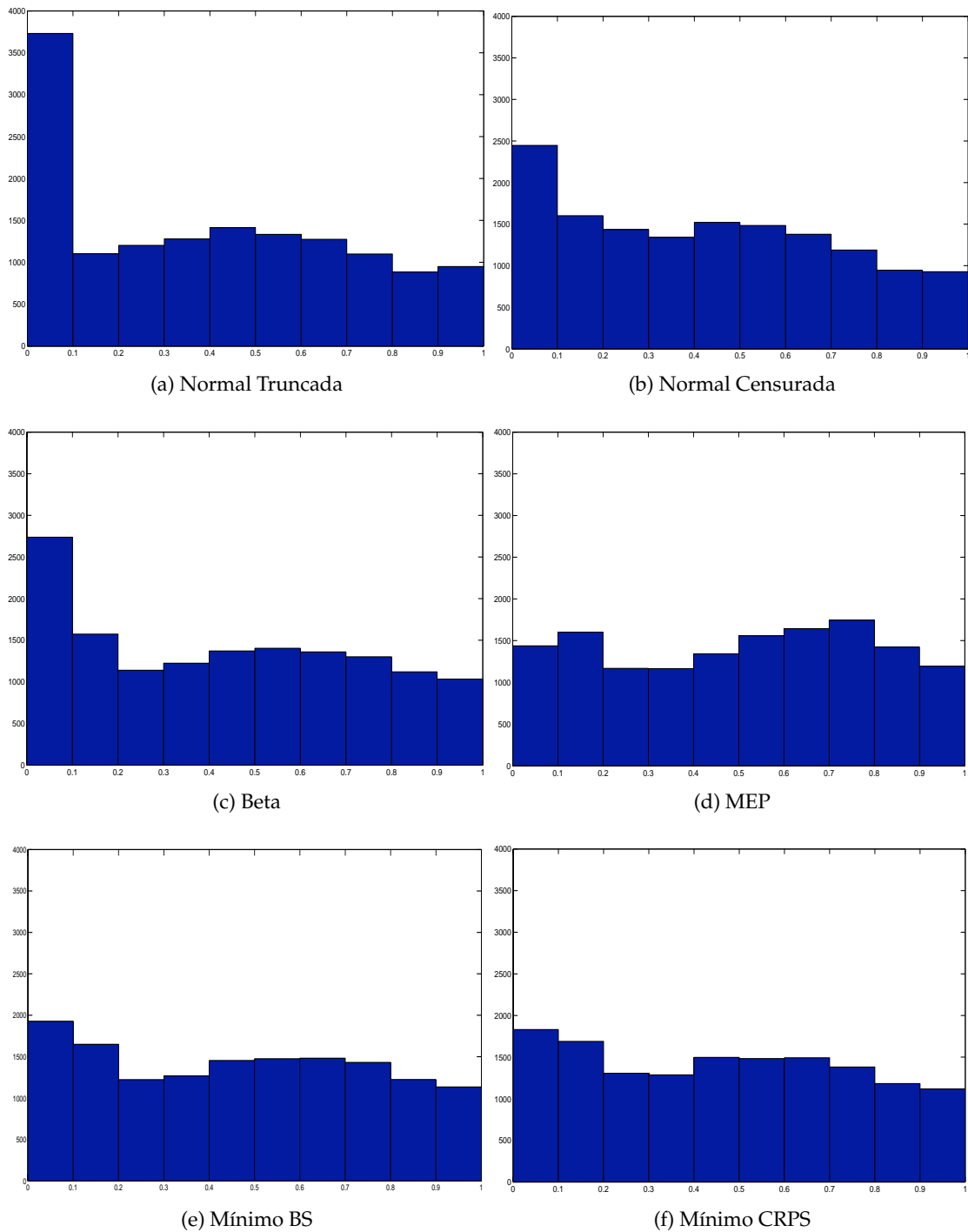


Figura 4.11: Histograma de los valores PIT obtenidos para diferentes densidades predictivas estimadas para un parque eólico a horizonte  $h = 1$ .

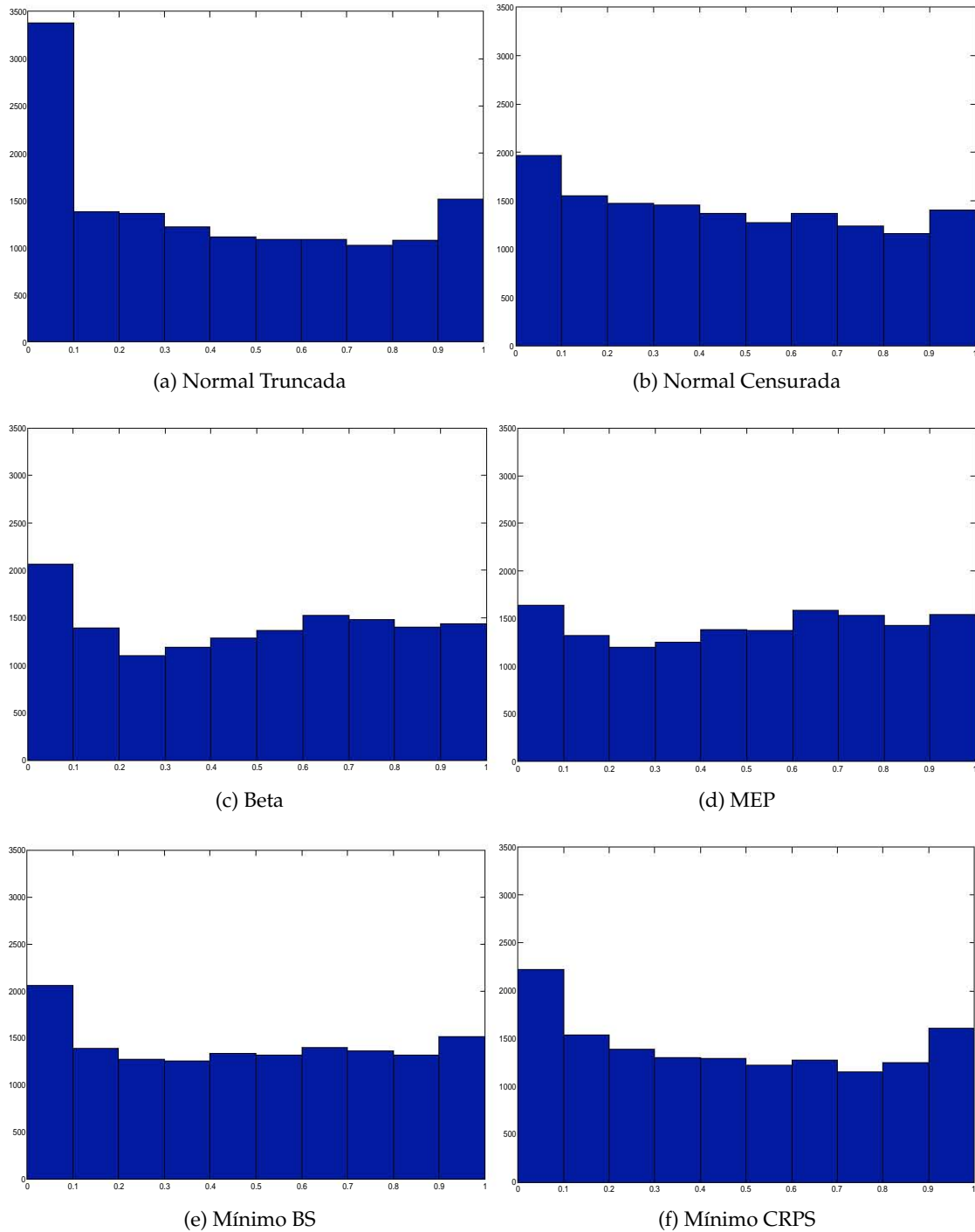


Figura 4.12: Histograma de los valores PIT obtenidos para diferentes densidades predictivas estimadas para un parque eólico a horizonte  $h = 6$ .

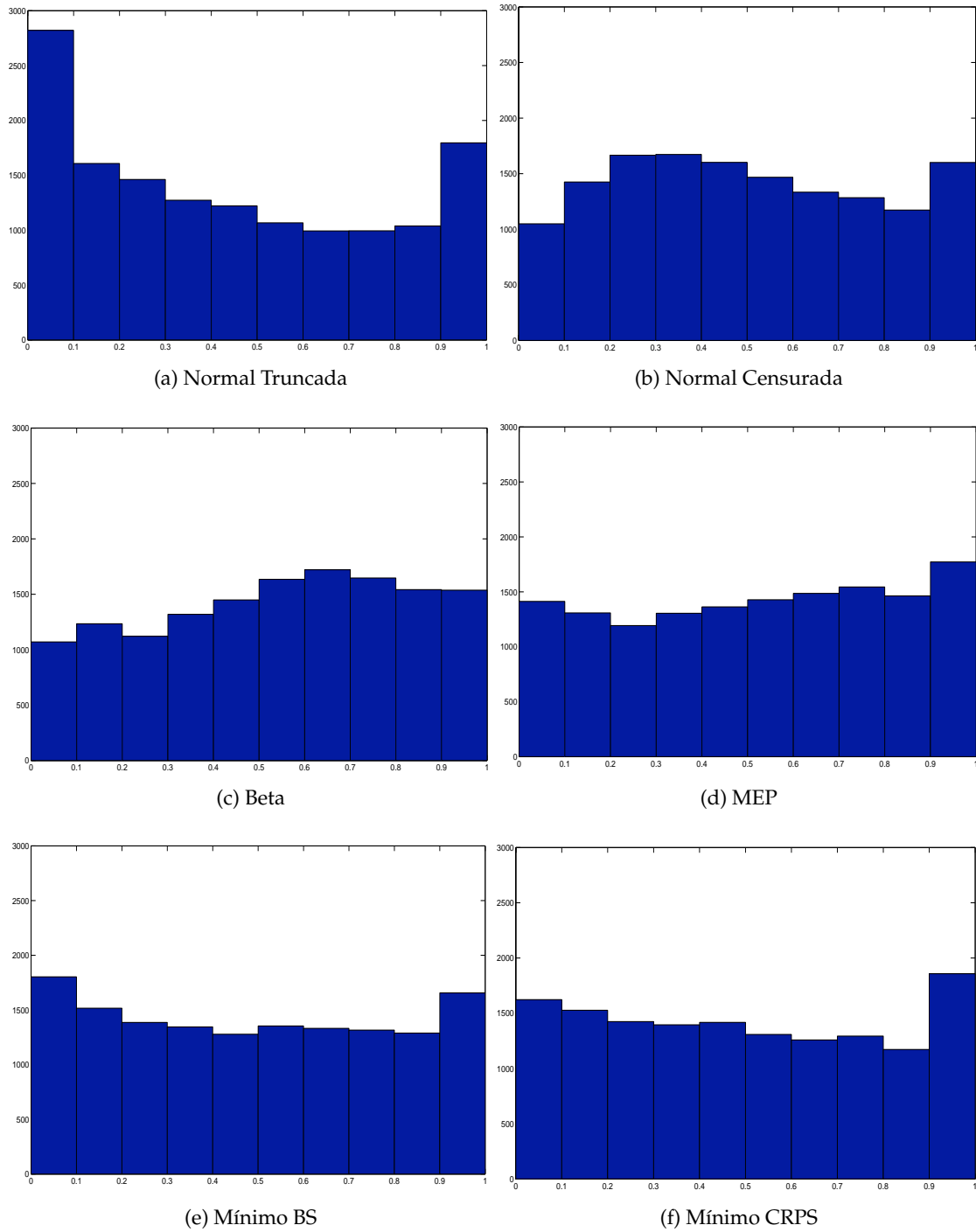


Figura 4.13: Histograma de los valores PIT obtenidos para diferentes densidades pre-  
dictivas estimadas para un parque eólico a horizonte  $h = 12$ .

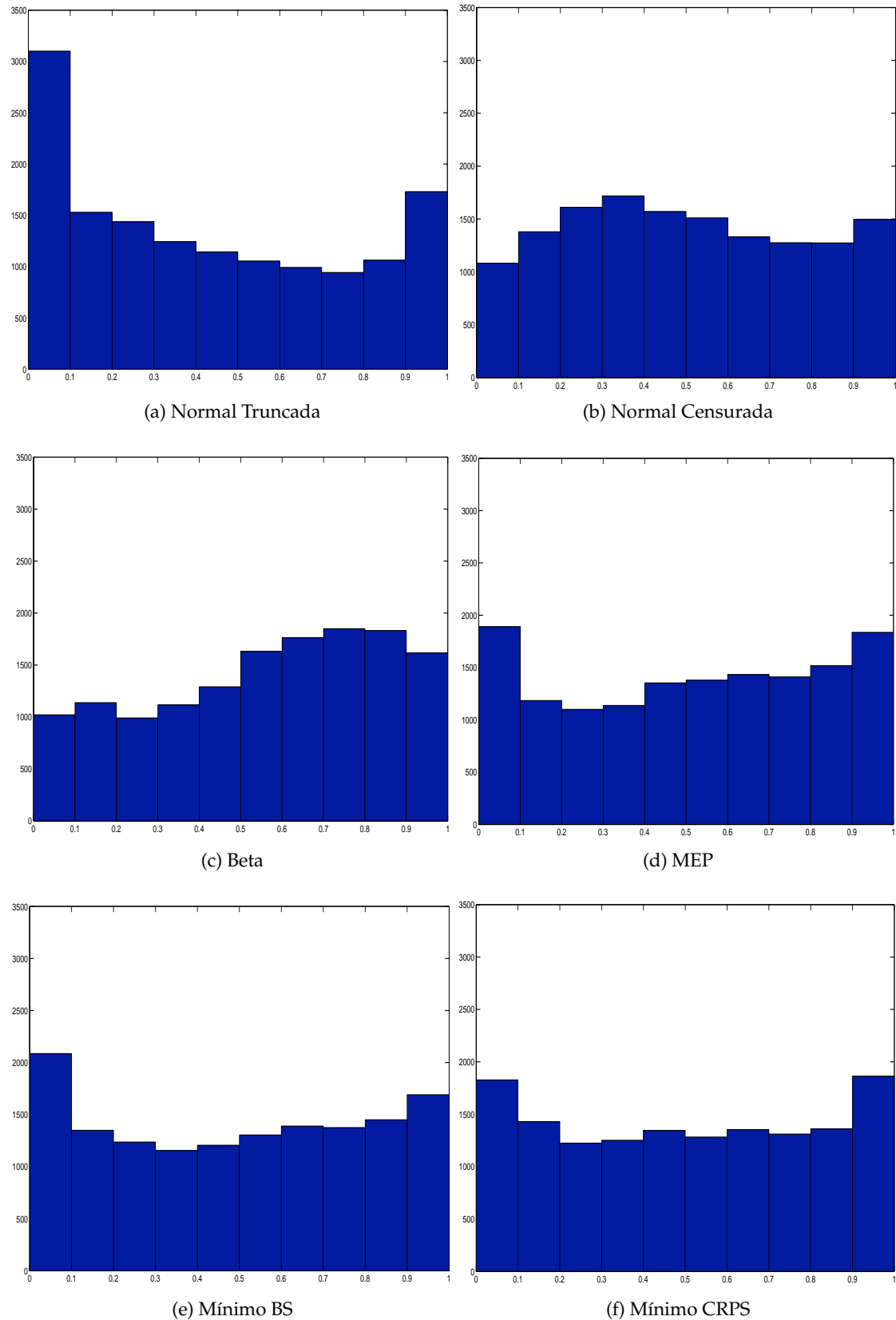


Figura 4.14: Histograma de los valores PIT obtenidos para diferentes densidades predictivas estimadas para un parque eólico a horizonte  $h = 24$ .

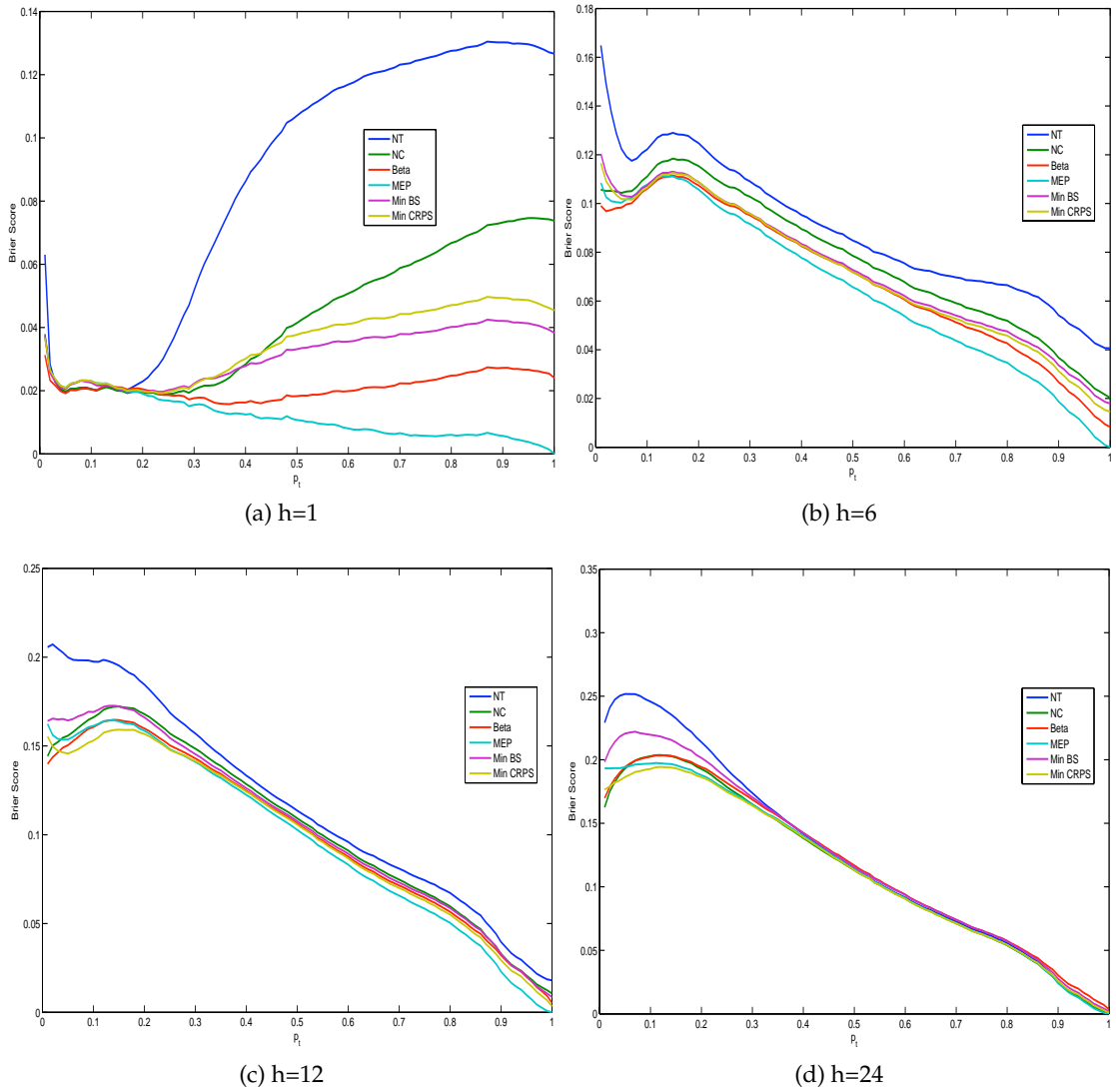


Figura 4.15: Valor del Brier Score obtenido para las diferentes distribuciones estimadas en un parque eólico empleando para diferentes horizontes de predicción  $h$ .

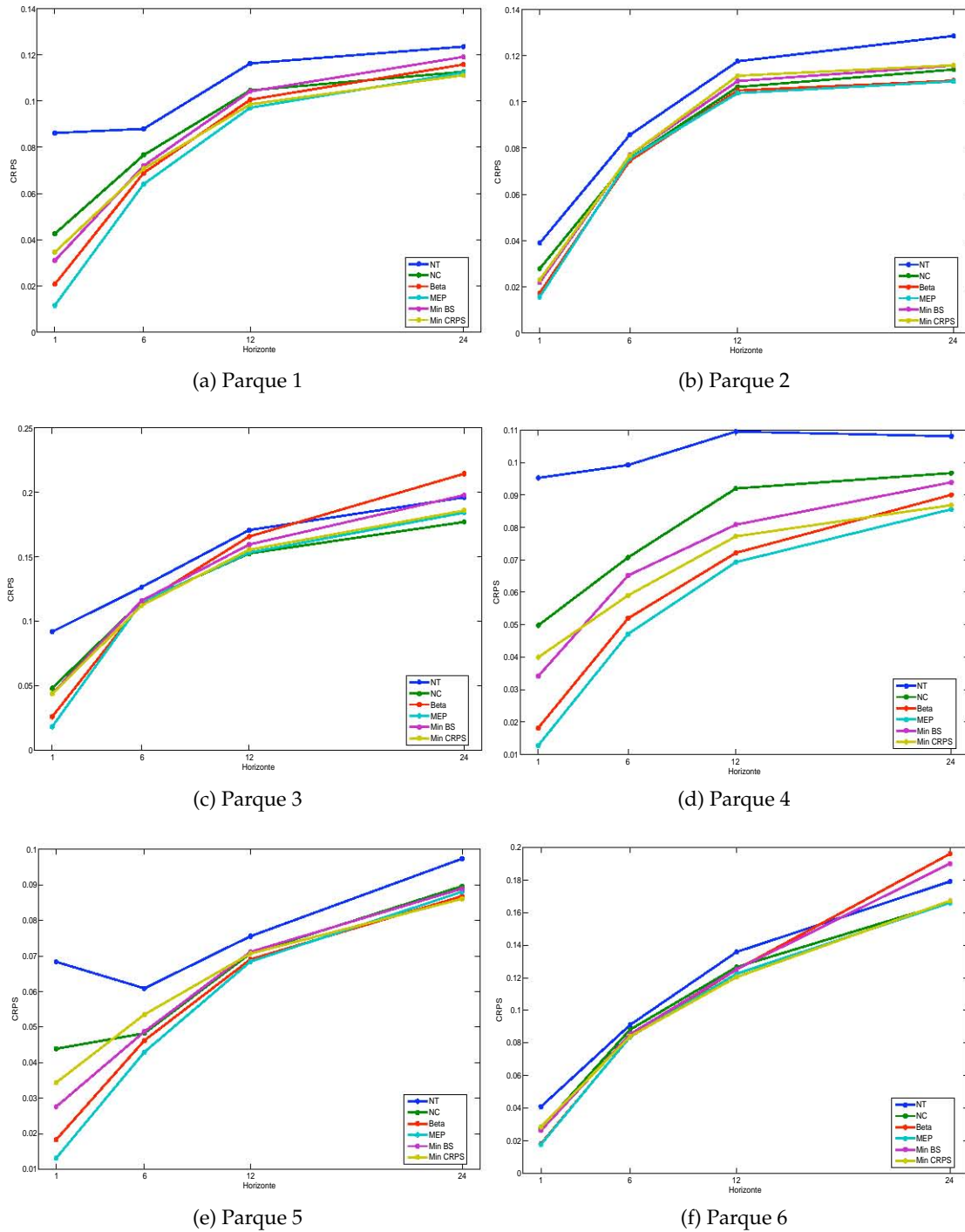


Figura 4.16: Valor del CRPS para las diferentes distribuciones predictivas empleadas en los diferentes parques eólicos disponibles.



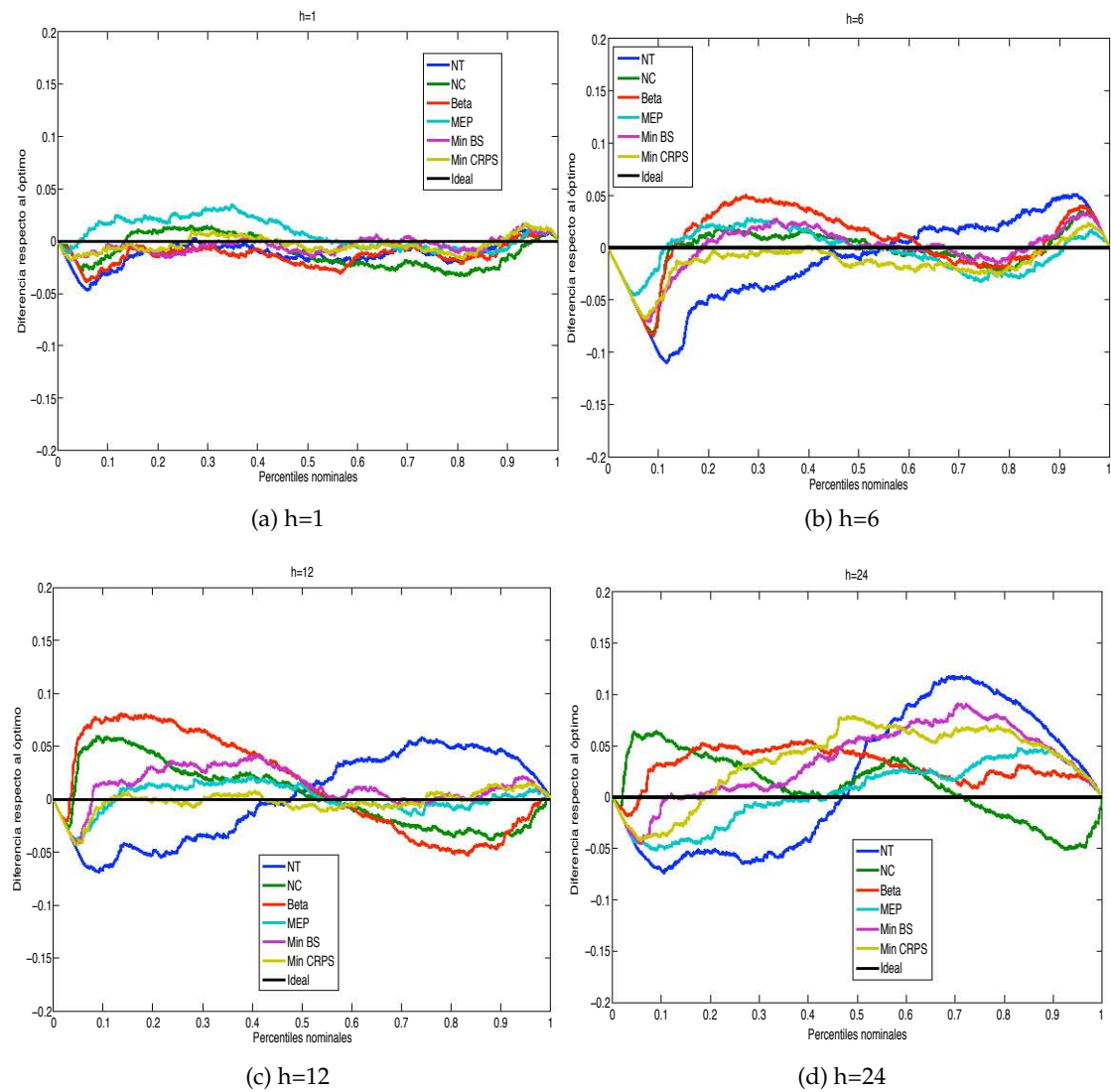


Figura 4.17: Comparación entre los percentiles obtenidos al evaluar la senda de densidades predictivas estimadas en 5 parques eólico empleando diferentes distribuciones



## Capítulo 5

# Predicción de producción eólica empleando modelos SETAR

La energía eólica tiene una clara desventaja respecto a otras fuentes de energía debido a su incertidumbre. Dada la naturaleza estocástica del viento que moverá el rotor del aerogenerador, no es posible conocer con exactitud la cantidad de potencia que se generará en los próximos periodos. Por ello, el cálculo de predicciones es una tarea crítica para poder gestionar esta energía y que pueda ser integrada en un sistema eléctrico. El horizonte de las predicciones dependerá de la tarea que estemos realizando. Por ejemplo, para programar labores de mantenimiento en un parque eólico serán necesarias predicciones de viento a varios días, mientras que para gestionar congestiones en el sistema eléctrico el horizonte será de horas o minutos. Asimismo, para participar en el mercado eléctrico diario, serán necesarias predicciones de hasta 48 horas de anticipación, mientras que la participación en mercados intradiarios precisan de predicciones de hasta 4 horas de anticipación.

En este capítulo se van a mostrar predicciones a muy corto plazo, puesto que serán predicciones a pocas horas. Este tipo de predicciones serán útiles para acudir a mercados intradiarios de energía así como para gestionar y controlar el funcionamiento del parque eólico.

### 5.1 Predicciones a corto plazo de potencia eólica

Existe una abundante bibliografía sobre modelos y métodos de predicción de potencia eólica, así como de la modelización y predicción de la velocidad del viento, dirección y demás factores que puedan afectar al rendimiento de un aerogenerador. Por ello solo haremos una breve revisión acerca de algunas de las propuestas representativas que

se han hecho empleando diferentes modelos de series temporales. En los estudios de Costa et al. (2008) y Giebel et al. (2011) podemos ver una amplia revisión histórica de las diferentes técnicas que se han ido proponiendo en los últimos 30 años.

Las diferentes propuestas de predicción empleando modelos de series temporales pueden ser separadas en dos tipos. Por una parte tenemos los modelos univariantes, y por otra los modelos de regresión dinámica en función de otra variable, que habitualmente será la velocidad prevista. Los primeros serán eficientes a horizontes de predicción bajos, mientras que cuando dicho horizonte aumente los modelos de regresión dinámica serán más adecuados.

### Modelos univariantes

Seguramente el primer trabajo que predijo la potencia eólica generada empleando un modelo univariante de series temporales fue el de Brown, Katz y Murphy (1984). Su idea fue realizar una transformación sobre los datos de la velocidad del viento con una frecuencia horaria mediante la raíz cuadrada para así obtener normalidad en los datos. Entonces ajustaban a los datos transformados un modelo autorregresivo y realizaban predicciones de la velocidad del viento, mediante las que extrapolaban los valores de la potencia empleando una curva de potencia estimada.

El modelo univariante clásico en el análisis de series temporales lineales es el modelo ARMA, gracias, especialmente, a la metodología desarrollada por Box y Jenkins (ver, Box y Jenkins, 1970; Pankratz, 1983) que permitió su modelización de manera sencilla. Fellows y Hill (1990) desarrollaron un procedimiento basado en dicha metodología para calcular predicciones de la velocidad del viento. Otras propuestas de modelos lineales ARMA aplicados a la energía eólica son, entre muchas otras, las de Tantareanu (1992), Kamal y Jafri (1997) y Torres et al. (2005).

Bossanyi (1985) propuso aplicar un filtro de Kalman para predecir la velocidad del viento con una frecuencia de 1 minuto empleando los 6 valores anteriores, en la que fue una de las primeras propuestas de modelos que emplean estimación recursiva. El uso de modelos no lineales, como ya explicamos en la sección 1.2.2 tiene el problema del cálculo de las predicciones, que suele ser complejo y computacionalmente costoso. Algunas propuestas de técnicas no lineales aplicadas a la energía eólica son Kariniotakis et al. (1997), Kariniotakis et al. (1999) y Baile et al. (2011), en las cuales se emplean técnicas como Wavelets o redes neuronales.

### Modelos de regresión dinámica

Las propuestas vistas hasta ahora modelizan o bien la potencia eólica o bien la velocidad del viento, pero siempre de manera univariante. Otra opción es emplear modelos para la potencia eólica con variables exógenas diferentes, como puede ser la velocidad o la dirección del viento. Algunas propuestas son Nielsen et al. (1999), Giebel et al. (2001), Landberg et al. (2002) o Sánchez (2006b).

En este capítulo vamos a emplear un modelo no lineal univariante sobre la potencia eólica generada, en particular, usaremos un modelo SETAR inspirado en el comportamiento mecánico del aerogenerador. Para la identificación de este modelo SETAR se utilizará el procedimiento desarrollado en el capítulo 3.3. Además del modelo SETAR para la esperanza condicionada, se utilizará un modelo adaptativo para la densidad predictiva basado en la metodología desarrollada en el capítulo 4.

## 5.2 Modelos SETAR en potencia eólica

La utilidad de los modelos con umbrales va más allá de su uso como función a trozos. Una de sus principales ventajas es que dependiendo de la complejidad de la estructura identificada, un modelo con umbrales puede servir para modelizar funciones no lineales más complejas. Por ejemplo, como mostraron Fan y Yao (2005, p.130), una función sinusoidal puede ser aproximada por un modelo SETAR con 4 regímenes. Es decir, el uso de modelos SETAR puede emplearse como aproximaciones de funciones no lineales más complejas.

En ocasiones el uso de estructuras con más de una variable umbral nos permitirá además poder interpretar el comportamiento de la serie a estudiar. Por ejemplo, Tiao y Tsay (1994) propusieron modelizar el producto nacional bruto de EEUU empleando un modelo SETAR con 4 regímenes, entre los que el proceso irá cambiando dependiendo del valor del primer y del segundo retardo. Es decir, emplea dos variables umbral. Para dichas variables el valor del umbral correspondiente es 0, por lo que, cada uno de los regímenes se pueden interpretar en términos de periodos de recesión o expansión económica.

Pinson et al. (2008), entre otros, muestran que la potencia eólica tiene diferentes comportamientos dependiendo de diferentes regímenes. En particular, muestran cómo la varianza de la potencia será más baja a niveles bajos y altos de producción, mientras que tomará un valor más elevado en niveles medios. Por ello proponen emplear diferentes modelos por umbrales con 3 regímenes, como son el SETAR, el STAR y el MSAR. El problema principal de su propuesta es que la estimación de los umbrales es realizada mediante un algoritmo de búsqueda intensiva, que seleccionará aquellos

umbrales que minimicen la suma de cuadrados residual de la estimación, dado que suponen que el número de parámetros AR en cada régimen es conocido. Además sólo realizan predicciones a horizonte  $h = 1$  para evitar el problema descrito en la sección 1.2.2.

La idea que ha motivado la modelización que aquí se propone es similar a la empleada por Tiao y Tsay (1994). El objetivo es comprobar si el comportamiento de la potencia eólica producida es igual cuando la producción crece y cuando decrece. Para ello, buscaremos la existencia de un modelo SETAR cuya variable umbral será el incremento del primer retardo  $\Delta p_{t-1} = p_{t-1} - p_{t-2}$ . Este comportamiento diferenciado entre los incrementos y los decrementos de potencia está justificado debido a la inercia de los componentes mecánicos del aerogenerador, así como a otros instrumentos de control del mismo que tendrán diferente comportamiento en las subidas y en las bajadas (ver, p.ej., Burton et al., 2001). Para comprobar la posible existencia de un modelo SETAR con esta estructura empleamos la herramienta ARLS descrita en 3.3.

En la figura 5.1 vemos el resultado obtenido en uno de los parques eólicos estudiados al aplicar la herramienta ARLS empleando como variable umbral  $\Delta p_{t-1}$ . En ese caso, el procedimiento Aut-ARLS seleccionó estimar un modelo AR(1) con parámetros variables y sin constante. En la figura vemos la secuencia de estimaciones del proceso ordenado respecto a  $\Delta p_{t-1}$ . Como vemos la herramienta ARLS identifica claramente un cambio de régimen en torno a 0, lo que confirma que la inercia de la potencia eólica tendrá un comportamiento diferente dependiendo de si está en un periodo de crecimiento o decrecimiento.

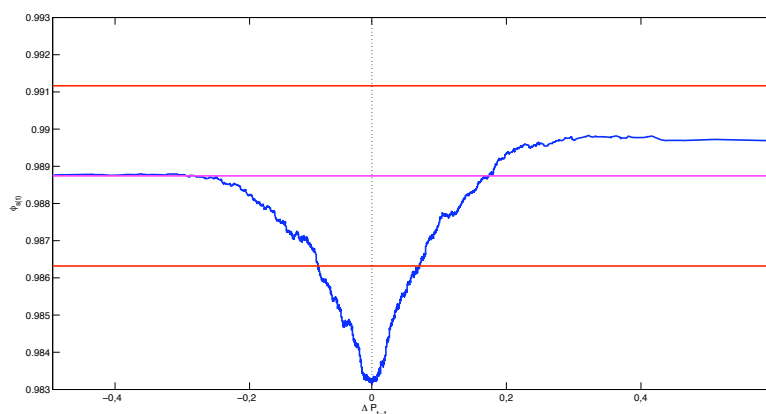


Figura 5.1: Estimación adaptativa de un modelo TV-AAR(1) sobre la potencia ordenada de acuerdo al incremento del primer retardo

Dado que hemos comprobado la existencia de dos tipos de comportamiento diferenciados, será interesante comprobar qué sucederá en cada uno de los casos. Lo primero,

por lo tanto, será comprobar qué sucede cuando la potencia generada decrece, es decir cuando  $\Delta p_{t-1} \leq 0$ . Aplicamos a dichos datos el procedimiento automático Aut-ARLS (ver sección 3.4), seleccionando el procedimiento como variable umbral el primer retardo de la potencia  $p_{t-1}$  y encontrando un modelo con 3 regímenes. Es decir, se identifican regímenes de comportamiento diferentes para niveles bajos, medios y altos de potencia. En la figura 5.2 podemos ver los umbrales detectados.

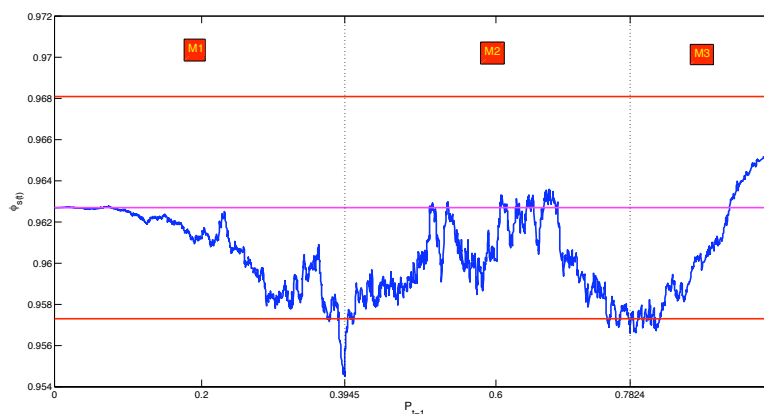


Figura 5.2: Estimación adaptativa de un modelo TV-AAR(1) sobre la potencia ordenada de acuerdo a su primer retardo.

Repetimos el mismo experimento para el caso en el que la potencia generada crece obteniendo resultados similares, solo que en esta ocasión, como podemos ver en la figura 5.3, Aut-ARLS necesita 2 pasos para encontrar los umbrales. Es decir, como explicamos en la sección 3.4 en realiza una estimación recursiva sobre el proceso ordenado respecto a su primer retardo. Como vemos en la figura 5.3a, se identifica un umbral cuando  $p_{t-1} = 0.6514$ . Entonces, el procedimiento automático, busca la posible existencia de más umbrales, seleccionando en el siguiente paso otro umbral cuando  $p_{t-1} = 0.2988$ , como vemos en la figura 5.3b.

Así pues en los datos provenientes del parque analizado se ha identificado un modelo SETAR con 6 regímenes de comportamiento, cuya estructura puede verse en la figura 5.4. En ella podemos ver los 6 regímenes de comportamiento, los 3 de la parte superior a la diagonal se corresponden a incrementos negativos de la potencia y los 3 de la parte inferior a los incrementos positivos.

### 5.2.1 Patrón común detectado

Una vez detectada una estructura en los datos provenientes de un parque eólico hemos de comprobar si dicha estructura puede ser extrapolada a datos provenientes de diferentes parques eólicos. Para ello hemos empleado datos de potencia eólica de distintas

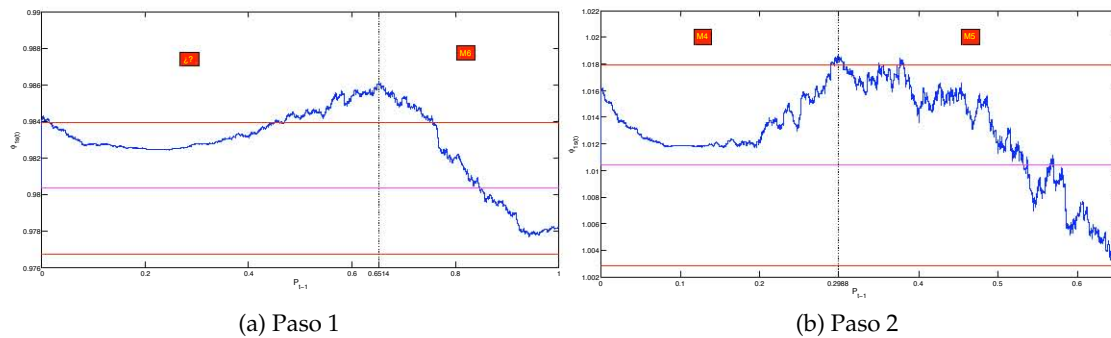


Figura 5.3: Estimación adaptativa de un modelo TV-AAR(1) sobre la potencia creciente ordenada de acuerdo a su primer retardo

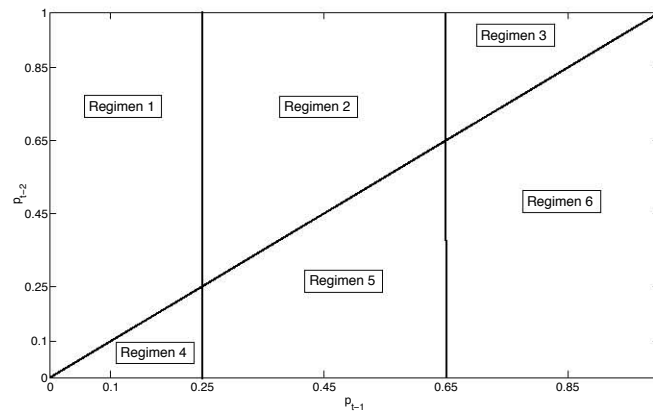


Figura 5.4: Estructura SETAR con 6 regímenes detectada en potencia eólica.

características:

- 36 ficheros de datos de potencia media producida por parques eólicos con una frecuencia horaria. 18 de los parques tienen un tamaño en torno a las 5000 observaciones, 16 de ellos en torno a las 15000 observaciones y 2 cerca de las 38000 observaciones.
- 19 ficheros de datos de potencia media generada por parques eólicos con una frecuencia de 15 minutos. El tamaño muestral de los ficheros está en torno a las 18000 observaciones.
- 21 ficheros de datos de potencia media generada por aerogeneradores individuales de un mismo parque eólico con una frecuencia de 10 minutos. El tamaño muestral de los datos supera las 51000 observaciones.



A todos los ficheros de datos descritos les aplicamos la herramienta Aut-ARLS para que busque la posible existencia del modelo SETAR con 6 regímenes descrito anteriormente. El resultado fue que en los 76 ficheros de datos se encontró un modelo SETAR con 6 regímenes. Además, en la figura 5.5 podemos ver los gráficos de caja de los umbrales detectados para todos los ficheros de datos. Como vemos, a pesar de emplear datos de diferentes frecuencias y tipos, se encuentra una estructura similar en todos los ejemplos.

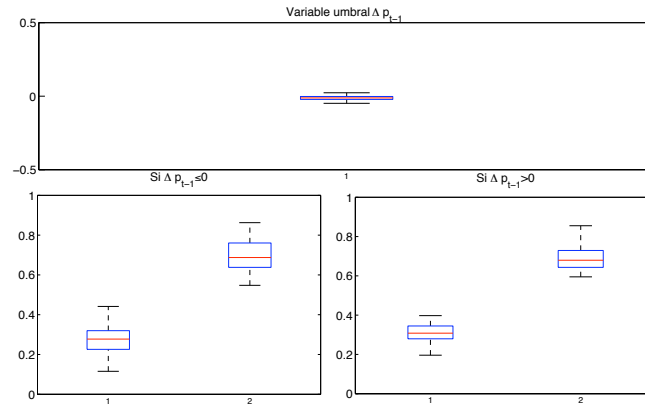


Figura 5.5: Gráficos de caja para los umbrales detectados en los 76 ficheros de datos disponibles

### 5.2.2 Modelos SETAR

A la vista de los resultados, proponemos modelizar la potencia eólica  $p_t$  producida mediante el modelo SETAR siguiente:

$$p_t = \begin{cases} \text{AR}(p_1) + e_t^{(1)} & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } p_{t-1} \leq r_L^1 \\ \text{AR}(p_2) + e_t^{(2)} & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } r_L^1 < p_{t-1} \leq r_U^1 \\ \text{AR}(p_3) + e_t^{(3)} & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } r_U^1 < p_{t-1} \\ \text{AR}(p_4) + e_t^{(4)} & , \text{ si } \Delta p_{t-1} > 0 \text{ y } p_{t-1} \leq r_L^2 \\ \text{AR}(p_5) + e_t^{(5)} & , \text{ si } \Delta p_{t-1} > 0 \text{ y } r_L^2 < p_{t-1} \leq r_U^2 \\ \text{AR}(p_6) + e_t^{(6)} & , \text{ si } \Delta p_{t-1} > 0 \text{ y } r_U^2 < p_{t-1} \end{cases}, \quad (5.1)$$

donde los valores  $r_L$  y  $r_U$  son calculados de manera automática por la herramienta Aut-ARLS para cada serie temporal concreta. Una vez seleccionados los ordenes autorregresivos  $p_1, \dots, p_6$  mediante el criterio propuesto por Galeano y Peña (2007), los parámetros serán estimados mediante el método de mínimos cuadrados descrito en la sección 2.3.1.

En general, los modelos estimados en datos provenientes de las mismas frecuencias

son muy semejantes en cuanto a la estructura autorregresiva estimada en cada régimen. En (5.2) podemos ver la estructura de un parque con frecuencia horaria. Como vemos el orden AR es 1 para todos los regímenes, y tan solo en alguno de los ejemplos aparece un segundo orden autorregresivo para alguno de los regímenes. Otra característica comúnmente observada es la falta de constante en los regímenes extremos, es decir, cuando el proceso es decreciente y su primer retardo tiene un valor próximo a cero, y cuando el proceso es creciente y su primer retardo toma un valor próximo a uno.

$$p_t = \begin{cases} 0.93p_{t-1} + e_t^{(1)} & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } p_{t-1} \leq 0.37 \\ 0.02 + 0.87p_{t-1} + e_t^{(2)} & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } 0.37 < p_{t-1} \leq 0.69 \\ -0.1 + 0.82p_{t-1} + e_t^{(3)} & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } 0.69 < p_{t-1} \\ 0.1 + 0.89p_{t-1} + e_t^{(4)} & , \text{ si } \Delta p_{t-1} > 0 \text{ y } p_{t-1} \leq 0.29 \\ 0.06 + 0.92p_{t-1} + e_t^{(5)} & , \text{ si } \Delta p_{t-1} > 0 \text{ y } 0.29 < p_{t-1} \leq 0.65 \\ 0.91p_{t-1} + e_t^{(6)} & , \text{ si } \Delta p_{t-1} > 0 \text{ y } 0.65 < p_{t-1} \end{cases} \quad (5.2)$$

En (5.3) vemos un ejemplo de modelo estimado en un parque con datos de frecuencia 15 minutos. Como vemos la principal diferencia al aumentar la frecuencia es el aumento del orden autorregresivo en los regímenes.

$$p_t = \begin{cases} 1.19p_{t-1} - 0.30p_{t-2} + e_t^{(1)} & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } p_{t-1} \leq 0.31 \\ 0.01 + 1.31p_{t-1} - 0.46p_{t-2} + 0.14p_{t-3} + e_t^{(2)} & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } 0.31 < p_{t-1} \leq 0.77 \\ 0.04 + 1.32p_{t-1} - 0.4830p_{t-2} + e_t^{(3)} & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } 0.77 < p_{t-1} \\ 1.59p_{t-1} - 0.69p_{t-2} + e_t^{(4)} & , \text{ si } \Delta p_{t-1} > 0 \text{ y } p_{t-1} \leq 0.26 \\ 0.03 + 1.25p_{t-1} - 0.33p_{t-2} + e_t^{(5)} & , \text{ si } \Delta p_{t-1} > 0 \text{ y } 0.26 < p_{t-1} \leq 0.77 \\ 1.11p_{t-1} - 0.26p_{t-2} + 0.11p_{t-3} + e_t^{(6)} & , \text{ si } \Delta p_{t-1} > 0 \text{ y } 0.77 < p_{t-1} \end{cases} \quad (5.3)$$

### 5.2.3 Modelos SETAR con densidad condicionada variable en el tiempo

Los modelos SETAR permiten que la media y la varianza del proceso cambien en el tiempo dependiendo de los diferentes regímenes, pero dentro de los regímenes permanecerán constantes. En ocasiones, como mostramos en la sección 1.1.5, existe la posibilidad de crear un modelo que permita que la varianza cambie dentro de los regímenes, empleando un modelo ARCH para la varianza condicionada. En esta sección vamos a proponer emplear un modelo SETAR en el que el término de error irá cambiando de forma similar a la mostrada en el capítulo 4. Es decir, dependiendo de un polinomio de la potencia prevista cuyos parámetros cambiarán a lo largo del tiempo. De esta manera, permitiremos que los momentos condicionados de los residuos cambien a lo largo

del tiempo sin depender exclusivamente de los regímenes. A este modelo lo llamaremos SETAR-TVCD (*Self Exciting Threshold Autoregressive with Time Varying Conditional Density*), y vendrá definido por

$$p_t = m(p_{t-1}, \dots, p_{t-h}) + \sigma(p_{t-1}, \dots, p_{t-h}), \quad (5.4)$$

donde  $m(\cdot)$  es una función a trozos lineal con forma

$$m(p_{t-1}, \dots, p_{t-h}) = \begin{cases} \text{AR}(p_1) & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } p_{t-1} \leq r_L^1 \\ \text{AR}(p_2) & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } r_L^1 < p_{t-1} \leq r_U^1 \\ \text{AR}(p_3) & , \text{ si } \Delta p_{t-1} \leq 0 \text{ y } r_U^1 < p_{t-1} \\ \text{AR}(p_4) & , \text{ si } \Delta p_{t-1} > 0 \text{ y } p_{t-1} \leq r_L^2 \\ \text{AR}(p_5) & , \text{ si } \Delta p_{t-1} > 0 \text{ y } r_L^2 < p_{t-1} \leq r_U^2 \\ \text{AR}(p_6) & , \text{ si } \Delta p_{t-1} > 0 \text{ y } r_U^2 < p_{t-1} \end{cases}, \quad (5.5)$$

y  $\sigma(\cdot) = E[e_t^2]$  sigue un polinomio definido por

$$\sigma(p_{t-1}, \dots, p_{t-k}) = \alpha_{0,t} + \alpha_{1,t}\hat{p}_t + \alpha_{2,t}\hat{p}_t^2 + \alpha_{3,t}\hat{p}_t^3 + v_t, \quad (5.6)$$

donde los valores  $\hat{p}_t$  serán calculados empleando el modelo autorregresivo de  $m(p_{t-1}, p_{t-2})$  correspondiente al régimen al que pertenezca la observación. De esta manera la estimación de (5.4) será realizada en dos etapas, primero estimaremos la media condicionada (5.6) dependiendo del régimen al que pertenezca el proceso, para continuación estimar de manera recursiva (5.6). La estimación recursiva se hará empleando el método de mínimos cuadrados adaptativos explicado en la sección 3.2.

### 5.3 Predicción empleando modelos SETAR-TVCD

Como dijimos en el capítulo 2.6 realizar predicciones empleando modelos SETAR será complicado en la mayoría de los casos. De las diferentes opciones detalladas en dicha sección, en esta tesis se ha empleado un procedimiento de Monte Carlo. Dicho procedimiento se basa en la simulación de diferentes trayectorias que el proceso  $\{p_t\}$  podría seguir al pasar del periodo  $t+h$  al  $t+h+1$  a partir de la distribución de las predicciones del horizonte  $h$ .

Sea  $X_t = \{p_{t-1}, p_{t-2}\}$  la variable umbral del modelo SETAR que empleamos para calcular las predicciones. Si queremos calcular predicciones a horizonte  $h = 1$ , la variable umbral  $X_t$  es conocida, por lo que la predicción  $\hat{p}_{t+1|t}$  puede ser calculada directamente empleando el modelo autorregresivo correspondiente al umbral del modelo SETAR (5.5) al que pertenezca.

El problema surge al calcular predicciones a horizonte  $h = 2$ , puesto que la variable umbral  $X_t$  sólo es observada en parte, así que no podremos asegurar a que umbral pertenece la predicción  $\hat{p}_{t+2|t}$ . Por ello, a partir de este instante de tiempo emplearemos un procedimiento de Monte Carlo para simular diferentes trayectorias posibles que el proceso podría tomar al pasar de  $t + 1$  a  $t + 2$ .

Para simular las diferentes trayectorias vamos a emplear el método de estimación de densidades predictivas mostrado en el capítulo 4. De esta manera, empleamos la predicción  $\hat{p}_{t+1|t}$  para estimar el  $j$ -ésimo momento condicionado del error de predicción mediante

$$m_{t+1|t}^{(j)} = E_{t+1} \left[ e_{t+1|t}^j \right], \quad j = 1, \dots, M, \quad (5.7)$$

donde la potencia del término de error se ajusta mediante el polinomio

$$e_{t+1|t}^j = \alpha_{0t+1}^{(j)} + \alpha_{1t+1}^{(j)} \hat{p}_{t+1|t} + \alpha_{2t+1}^{(j)} \hat{p}_{t+1|t}^2 + \alpha_{3t+1}^{(j)} \hat{p}_{t+1|t}^3 + v_{t+1}, \quad j = 1, \dots, M. \quad (5.8)$$

Estimar el modelo (5.8) no es posible, puesto que los errores de predicción no serán observados, ya que estamos realizando predicciones fuera de la muestra. Debido a ello se emplearán los últimos estimadores de los coeficientes de (5.8) dentro de la muestra, es decir, aquellos valores estimados en el instante de tiempo  $t$  condicionados a la información contenida en el instante de tiempo  $t - 1$ ,

$$e_{t|t-1}^j = \hat{\alpha}_{0t|t-1}^{(j)} + \hat{\alpha}_{1t|t-1}^{(j)} \hat{p}_{t|t-1} + \hat{\alpha}_{2t|t-1}^{(j)} \hat{p}_{t|t-1}^2 + \hat{\alpha}_{3t|t-1}^{(j)} \hat{p}_{t|t-1}^3 + v_t. \quad (5.9)$$

De esta manera, emplearemos los valores  $\hat{\alpha}_{t|t-1}^{(j)}$  junto con la predicción  $\hat{p}_{t+1|t}$  para estimar los momentos condicionados del error de predicción a horizonte  $h = 1$ , es decir, la estimación será

$$m_{t+1|t}^{(j)} = \hat{\alpha}_{0t|t-1}^{(j)} + \hat{\alpha}_{1t|t-1}^{(j)} \hat{p}_{t+1|t} + \hat{\alpha}_{2t|t-1}^{(j)} \hat{p}_{t+1|t}^2 + \hat{\alpha}_{3t|t-1}^{(j)} \hat{p}_{t+1|t}^3, \quad j = 1, \dots, M. \quad (5.10)$$

Una vez estimados los momentos condicionados podemos estimar como densidad predictiva  $f_{t+1}$  de la predicción  $\hat{p}_{t+1|t}$  las diferentes distribuciones mostradas en el capítulo 4, es decir, la Normal truncada, la Normal censurada, la Beta y la de Máxima Entropía. Empleando esta distribución generamos  $B$  números aleatorios  $\hat{p}_{t+1|t}^{(i)} \sim f_{t+1}$ , para  $i = 1, \dots, B$ , desde los que se estiman  $B$  trayectorias posibles  $\hat{p}_{t+2|t}^{(i)}$ , mediante el modelo SETAR (5.5) y empleando como variable umbral  $X_{t+1} = \left\{ \hat{p}_{t+1|t}^{(i)}, p_t \right\}$ . Empleando dichas trayectorias se calcula la predicción puntual, empleando alguna medida de centralidad. En esta tesis se han calculado mediante la media y la mediana de las trayectorias, es decir, se obtienen dos predicciones puntuales  $\hat{p}_{t+2|t}^{media}$  y  $\hat{p}_{t+2|t}^{mediana}$ .

Para calcular predicciones a horizontes más altos se repite el mismo proceso que

para horizonte  $h = 2$ . De esta manera, la predicción  $\hat{p}_{t+h|t}$  dado que la predicción  $\hat{p}_{t+h-1|t}$  ha sido previamente calculada, se estima empleando  $B$  trayectorias posibles que  $\hat{p}_{t+h-1|t}^{(i)}$  puede tomar. Para generar estos  $B$  valores se necesita la densidad predictiva  $f_{t+h-1}$ . Como ya dijimos, dicha densidad predictiva se estima empleando los momentos condicionados  $m_{t+h-1|t}^{(j)}$  que a su vez serán calculados mediante

$$m_{t+h-1|t}^{(j)} = \hat{\alpha}_{0t|t-h+1}^{(j)} + \hat{\alpha}_{1t|t-h+1}^{(j)} \hat{p}_{t+h-1|t} + \hat{\alpha}_{2t|t-h+1}^{(j)} \hat{p}_{t+h-1|t}^2 + \hat{\alpha}_{3t|t-h+1}^{(j)} \hat{p}_{t+h-1|t}^3, \quad j = 1, \dots, M, \quad (5.11)$$

donde los valores  $\hat{\alpha}_{t|t-h+1}^{(j)}$  son la última estimación obtenida dentro de la muestra a horizonte  $h - 1$ . Así, empleando los momentos condicionados  $m_{t+h-1|t}^{(j)}$  se estima la densidad predictiva  $f_{t+h-1}$  y se generan  $B$  posibles trayectorias  $\hat{p}_{t+h-1|t}^{(i)}$ , desde las que empleando el modelo SETAR se pueden calcular los  $B$  valores  $\hat{p}_{t+h|t}^{(i)}$  empleados para calcular la predicción puntual mediante su media o su mediana.

## 5.4 Predicciones puntuales de producción eólica

En esta sección se comprueba la eficacia de los modelos SETAR propuestos en las secciones 5.2.2 y 5.2.3 realizando predicciones sobre los diferentes datos de potencia eólica descritos en la sección 5.2.1. Compararemos los modelos SETAR entre sí, empleando las diferentes distribuciones mostradas en el capítulo 4. Además, dichos modelos SETAR son comparados con el mejor modelo lineal ARMA encontrado para cada fichero de datos. Para ello, se ha empleado un procedimiento de búsqueda del orden óptimo del modelo ARMA basado en la metodología Box-Jenkins, es decir, asumiendo que el proceso es estacionario, se seleccionan los órdenes máximos que el modelo ARMA puede tomar empleando la función de autocorrelación simple y parcial. A continuación se estiman las diferentes combinaciones posibles empleando dichos órdenes máximos y se selecciona aquél modelo que minimiza el criterio de selección AIC y el BIC. Por último hemos empleado un modelo TV-AR con parámetros variables definido por

$$p_t = \phi_{0t} + \phi_{1t}p_{t-1} + \dots + \phi_{pt}p_{t-p} + e_t, \quad (5.12)$$

donde la estimación de los parámetros variables se realiza empleando el método de mínimos cuadrados adaptativos descrito en la sección 3.2. Como factores de olvido se empleó el factor de olvido variable basado en la distancia de Cook (ver sección 3.2) y varios factores de olvido constantes, obteniendo mejores resultados con el factor de olvido variable.

Las predicciones en los modelos SETAR-TVCD han sido calculadas empleando el procedimiento de Monte Carlo anteriormente descrito empleando  $B = 500$  trayectorias. Para las predicciones mediante el modelo SETAR con error homocedástico en cada umbral se ha empleado un procedimiento de Monte Carlo similar al mostrado pero la estimación de los momentos se ha realizado mediante la estimación muestral de los momentos de los residuos dentro de cada régimen.

La muestra fue dividida en una submuestra de estimación y en otra de validación sobre la que se realizaron las predicciones fuera de la muestra. El tamaño de la muestra de estimación es fijado en el 75% de la muestra, por lo que el número de predicciones puntuales calculadas, dado el tamaño habitual de los ficheros de potencia eólica disponibles, supera las 2000. Las medidas calculadas para comparar los resultados son la raíz del error cuadrático medio (RECM) y el error absoluto medio (EAM).

En las tablas 5.1 y 5.3 vemos la media de los resultados para los 31 parques eólicos con datos horarios. Además, para facilitar la comparación entre modelos las tablas 5.2 y 5.4 muestran las medidas de error relativas, que resultan de dividir cada medida de error por la obtenida con el modelo SETAR-TVCD con distribución de máxima entropía. Para la tabla del error cuadrático medio se usa como referencia la media de las trayectorias, mientras que para el error medio absoluto se usa la mediana de las trayectorias. Es decir, siempre que el valor mostrado sea mayor o igual a 1, el modelo SETAR-TVCD empleando el MEP (empleando la media o la mediana) será la mejor propuesta. Como vemos en 5.2 en solo dos modelos SETAR empleando como distribución el MEP para un horizonte  $h = 2$  mejoran levemente al modelo SETAR-TVCD. Pero en general, los resultados son mejores para el SETAR-TVCD con el MEP, mejorando por ejemplo para un horizonte  $h = 6$  un 26% el valor obtenido con el TV-AR y un 69% al modelo lineal. En la tabla 5.4 vemos cómo los resultados se igualan más, pero aun así todos los valores son mayores o iguales a 1. Aunque a horizontes 5 y 6, los valores obtenidos con el SETAR-TVCD y con el TV-AR son prácticamente iguales.

Un comportamiento similar aunque con menos mejora para los modelos SETAR puede observarse en datos con frecuencia de 15 minutos, como puede verse en las tablas 5.5 y 5.7. Vuelve a no existir ningún valor menor a 1, con lo que el modelo SETAR-TVCD siempre es mejor, aunque en ocasiones estará muy parejo a los valores obtenidos con el modelo TV-AR, especialmente a horizontes altos y empleando el error medio absoluto. Resultados similares se obtienen para datos provenientes de aerogeneradores con frecuencia de 10 minutos, como puede verse en las tablas 5.9 y 5.11.

## 5.5 Conclusiones

Los resultados obtenidos mediante el cálculo de predicciones en 72 ficheros de potencia producida tanto por parques eólicos como por diferentes aerogeneradores de un mismo parque y con diferentes frecuencias de tiempo, muestran que el modelo SETAR-TVCD propuesto es una buena alternativa para predecir la producción eólica.

Esos mismos resultados obtenidos en datos de tan diferente procedencia nos confirman que la estructura de 6 regímenes detectada por la herramienta Aut-ARLS es adecuada para todo tipo de datos de producción eólica, sin importar la frecuencia o si son datos sobre la producción agregada de un parque o de la potencia desarrollada por aerogeneradores particulares.

Además, confirmando lo observado en el capítulo 4, la distribución de Máxima Entropía basada en la estimación adaptativa de los primeros cuatro momentos condicionales mejora las restantes distribuciones paramétricas empleadas en este trabajo. Las posibles líneas de investigación futuras se centrarán en la mejora del modelo SETAR-TVCD propuesto. Una posible mejora podría obtenerse encontrando un criterio de selección de momentos para la distribución de máxima entropía. Otra mejora posible es la estimación de la media condicionada empleando un factor de olvido próximo a 1, o el uso de otro tipo de modelos como podría ser un modelo MSAR. También será interesante realizar una evaluación de la densidad condicionada estimada.

Por último, posibles aplicaciones futuras sería probar si el patrón aquí detectado aparece en otro tipo de datos como pueden ser parques eólicos marítimos (datos *off-shore*) o en datos provenientes de los nuevos superaerogeneradores de más de 5 GW.

Media de la raíz del error cuadrático medio														
SETAR					SETAR-TVCD									
h	ARMA	TV-AR	Media			Mediana			Media			Mediana		
			NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP
1	,0899	,0873	,0836	,0836	,0836	,0836	,0836	,0836	,0835	,0835	,0835	,0835	,0835	,0835
2	,1423	,1336	,1378	,1529	,1567	,1151	,1240	,1376	,1446	,1154	,1294	,1290	,1223	,1293
3	,1836	,1639	,1795	,2040	,2035	,1538	,1541	,1746	,1826	,1540	,1594	,1582	,1394	,1589
4	,2279	,1876	,2102	,2399	,2387	,1747	,1732	,1994	,2069	,1756	,1821	,1804	,1552	,1813
5	,2672	,2069	,2330	,2659	,2639	,1950	,1853	,2151	,2202	,1885	,1995	,1980	,1702	,1988
6	,2994	,2229	,2504	,2860	,2837	,2090	,1935	,2243	,2297	,2030	,2133	,2121	,1771	,2128

Tabla 5.1: Media de la raíz del error cuadrático medio para los 31 parques eólicos horarios.

Valor relativo de la raíz del error cuadrático medio														
SETAR					SETAR-TVCD									
h	ARMA	TV-AR	Media			Mediana			Media			Mediana		
			NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP
1	1,08	1,05	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2	1,16	1,09	1,13	1,25	1,28	0,94	1,01	1,13	1,18	0,94	1,06	1,05	1,05	1,07
3	1,32	1,18	1,29	1,46	1,46	1,10	1,11	1,25	1,31	1,10	1,14	1,13	1,13	1,16
4	1,47	1,21	1,35	1,55	1,54	1,13	1,12	1,28	1,33	1,13	1,17	1,16	1,16	1,21
5	1,57	1,22	1,37	1,56	1,55	1,15	1,09	1,26	1,29	1,11	1,17	1,16	1,16	1,23
6	1,69	1,26	1,41	1,61	1,60	1,18	1,09	1,27	1,30	1,15	1,20	1,20	1,20	1,29

Tabla 5.2: Valor relativo de la raíz del error cuadrático medio obtenido para los 31 parques eólicos horarios.



Media del error medio absoluto																
SETAR										SETAR-TVCD						
Media										Mediana						
Mediana										Media						
Mediana										Mediana						
h	ARMA	TV-AR	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	MEP
1	,0502	,0519	,0490	,0490	,0490	,0490	,0490	,0490	,0490	,0490	,0485	,0485	,0485	,0485	,0485	,0485
2	,0785	,0772	,0933	,0919	,0918	,0853	,0853	,0853	,0849	,0834	,0860	,0812	,0787	,0816	,0842	,0749
3	,0993	,0887	,1266	,1244	,1238	,1011	,1115	,1117	,1108	,0975	,1125	,1042	,0992	,0976	,1091	,0936
4	,1171	,1092	,1532	,1524	,1515	,1141	,1357	,1357	,1342	,1086	,1320	,1220	,1154	,1104	,1279	,1050
5	,1354	,1161	,1750	,1770	,1756	,1261	,1578	,1584	,1561	,1191	,1468	,1362	,1292	,1225	,1425	,1208
6	,1555	,1209	,1932	,1980	,1963	,1311	,1791	,1800	,1771	,1244	,1588	,1485	,1417	,1281	,1543	,1209

Tabla 5.3: Media del error absoluto medio para los 31 parques eólicos horarios.

Valor relativo del error medio absoluto																
SETAR										SETAR-TVCD						
Media										Mediana						
Mediana										Media						
Mediana										Mediana						
h	ARMA	TV-AR	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	MEP
1	1,04	1,07	1,01	1,01	1,01	1,01	1,01	1,01	1,01	1,01	1,00	1,00	1,00	1,00	1,00	1,00
2	1,12	1,10	1,33	1,31	1,31	1,21	1,21	1,21	1,21	1,19	1,23	1,16	1,12	1,16	1,20	1,07
3	1,19	1,06	1,51	1,49	1,48	1,21	1,33	1,34	1,32	1,17	1,35	1,25	1,19	1,17	1,30	1,12
4	1,12	1,04	1,46	1,45	1,44	1,09	1,29	1,29	1,28	1,03	1,26	1,16	1,10	1,05	1,22	1,03
5	1,17	1,00	1,51	1,53	1,52	1,09	1,36	1,37	1,35	1,03	1,27	1,18	1,12	1,06	1,23	1,05
6	1,29	1,00	1,60	1,64	1,62	1,08	1,48	1,49	1,46	1,03	1,31	1,23	1,17	1,06	1,28	1,09

Tabla 5.4: Valor relativo del error medio absoluto obtenido para los 31 parques eólicos horarios.

		Media de la raíz del error cuadrático medio													
		SETAR							SETAR-TVCD						
		Media				Mediana			Media				Mediana		
h	ARMA	TV-AR	NT	NC	BETA	MEP	MEP	NT	NC	BETA	MEP	MEP	NT	NC	BETA
1	,0598	,0576	,0506	,0506	,0506	,0506	,0506	,0501	,0501	,0501	,0501	,0501	,0501	,0501	,0501
2	,0938	,0809	,0876	,0877	,0877	,0855	,0858	,0858	,0863	,0869	,0860	,0860	,0787	,0789	,0754
3	,1183	,1007	,1150	,1153	,1152	,1129	,1110	,1110	,1121	,1133	,1127	,1127	,0985	,0994	,0935
4	,1385	,1177	,1387	,1393	,1392	,1388	,1316	,1316	,1329	,1346	,1330	,1330	,1141	,1142	,1078
8	,1985	,1601	,2143	,2170	,2165	,2104	,2020	,2020	,2042	,2086	,2077	,2077	,1578	,1577	,1574
12	,2405	,1879	,2695	,2759	,2747	,2675	,2488	,2488	,2527	,2583	,2582	,2582	,1846	,1849	,1737
16	,2798	,2104	,3073	,3181	,3158	,3159	,2828	,2828	,2884	,2942	,2940	,2940	,2062	,2062	,1892
20	,3152	,2284	,3338	,3459	,3433	,3497	,3070	,3070	,3142	,3191	,3189	,3189	,2233	,2233	,2070
24	,3497	,2445	,3562	,3698	,3668	,3719	,3265	,3265	,3357	,3408	,3405	,3405	,2393	,2392	,2209

Tabla 5.5: Media de la raíz del error cuadrático medio para los 19 parques eólicos 15 minutos.

		Valor relativo de la raíz del error cuadrático medio													
		SETAR							SETAR-TVCD						
		Media				Mediana			Media				Mediana		
h	ARMA	TV-AR	NT	NC	BETA	MEP	MEP	NT	NC	BETA	MEP	MEP	NT	NC	BETA
1	1,19	1,15	1,01	1,01	1,01	1,01	1,01	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2	1,34	1,15	1,25	1,25	1,25	1,22	1,22	1,12	1,12	1,12	1,00	1,00	1,12	1,12	1,08
3	1,30	1,10	1,26	1,26	1,26	1,24	1,24	1,08	1,08	1,08	1,00	1,00	1,08	1,08	1,03
4	1,29	1,10	1,30	1,30	1,30	1,30	1,23	1,07	1,07	1,07	1,00	1,00	1,07	1,07	1,01
8	1,37	1,11	1,48	1,50	1,50	1,46	1,40	1,09	1,09	1,09	1,00	1,00	1,09	1,09	1,09
12	1,44	1,12	1,61	1,65	1,64	1,60	1,48	1,10	1,10	1,10	1,00	1,00	1,10	1,11	1,04
16	1,49	1,12	1,63	1,69	1,68	1,68	1,50	1,10	1,10	1,10	1,00	1,00	1,10	1,11	1,01
20	1,55	1,12	1,64	1,70	1,69	1,72	1,51	1,10	1,10	1,10	1,00	1,00	1,10	1,11	1,02
24	1,59	1,11	1,62	1,68	1,66	1,69	1,48	1,09	1,09	1,09	1,00	1,00	1,09	1,10	1,00

Tabla 5.6: Valor relativo de la raíz del error cuadrático medio obtenido para los 19 parques eólicos 15 minutos.

		Media del error medio absoluto											
		SETAR						SETAR-TVCD					
h	ARMA	Media						Mediana					
		TV-AR	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA
1	,0760	,0711	,0671	,0671	,0671	,0671	,0671	,0671	,0671	,0671	,0658	,0658	,0658
2	,1295	,1151	,1232	,1227	,1243	,1228	,1148	,1147	,1159	,1144	,1192	,1173	,1109
3	,1680	,1485	,1651	,1643	,1656	,1607	,1506	,1509	,1527	,1439	,1590	,1558	,1445
4	,2013	,1767	,2020	,2010	,2020	,1938	,1805	,1812	,1838	,1687	,1916	,1877	,1733
8	,2933	,2557	,2981	,2968	,2976	,2785	,2662	,2679	,2725	,2377	,2777	,2529	,2482
12	,3578	,3273	,3667	,3654	,3658	,3410	,3285	,3305	,3377	,2904	,3352	,3289	,3054
16	,4176	,3417	,4340	,4329	,4330	,4017	,3860	,3885	,3982	,3413	,3901	,3833	,3573
20	,4747	,3806	,4954	,4948	,4944	,4544	,4408	,4451	,4564	,3854	,4376	,4298	,3994
24	,5449	,4143	,5760	,5805	,5781	,5181	,4952	,5040	,5167	,4370	,4826	,4733	,4541

Tabla 5.7: Media del error medio absoluto para los 19 parques eólicos 15 minutales.

		Valor relativo del error medio absoluto											
		SETAR						SETAR-TVCD					
h	ARMA	Media						Mediana					
		TV-AR	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA
1	1,16	1,08	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,00	1,00	1,00
2	1,16	1,03	1,10	1,10	1,11	1,10	1,03	1,03	1,04	1,02	1,07	1,05	0,98
3	1,16	1,02	1,14	1,13	1,14	1,11	1,04	1,04	1,05	0,99	1,10	1,08	0,99
4	1,17	1,02	1,17	1,16	1,17	1,12	1,04	1,05	1,06	0,98	1,11	1,09	1,00
8	1,18	1,03	1,20	1,20	1,20	1,12	1,07	1,08	1,10	0,96	1,12	1,10	1,00
12	1,21	1,11	1,24	1,24	1,24	1,15	1,11	1,12	1,14	0,98	1,13	1,11	1,03
16	1,22	1,00	1,27	1,26	1,26	1,17	1,13	1,13	1,16	1,00	1,14	1,12	1,04
20	1,24	1,00	1,30	1,29	1,29	1,19	1,15	1,16	1,19	1,01	1,14	1,12	1,05
24	1,32	1,00	1,39	1,40	1,40	1,25	1,20	1,22	1,25	1,06	1,17	1,14	1,10

Tabla 5.8: Valor relativo del error medio absoluto obtenido para los 19 parques eólicos 15 minutales.

		Media de la raíz del error cuadrático medio																	
		SETAR									SETAR-TVCD								
h		Media									Mediana								
		TV-AR	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP	MEP
1	ARMA	,0447	,0428	,0428	,0428	,0428	,0428	,0428	,0428	,0428	,0428	,0428	,0428	,0426	,0426	,0426	,0426	,0426	,0426
2	,0456	,0635	,0661	,0675	,0693	,0670	,0676	,0739	,0753	,0693	,0595	,0660	,0694	,0587	,0620	,0619	,0625	,0594	
3	,0688	,0763	,0846	,0847	,0847	,0826	,0872	,0849	,0909	,0830	,0761	,0760	,0760	,0677	,0716	,0714	,0722	,0698	
4	,0825	,0857	,0922	,0924	,0924	,0905	,0963	,0992	,0959	,0950	,0792	,0790	,0789	,0731	,0833	,0829	,0850	,0754	
5	,0919	,0934	,1002	,1006	,1005	,1003	,1027	,1089	,1075	,1074	,0827	,0824	,0823	,0773	,0903	,0897	,0935	,0810	
6	,1004	,0992	,1064	,1071	,1070	,1075	,1099	,1111	,1128	,1098	,0866	,0865	,0865	,0793	,0926	,0923	,0979	,0827	
12	,1048	,1326	,1403	,1382	,1411	,1381	,1456	,1426	,1457	,1426	,1118	,1119	,1120	,1015	,1185	,1178	,1228	,1068	
18	,1436	,1515	,1562	,1566	,1574	,1559	,1639	,1643	,1665	,1630	,1216	,1218	,1221	,1109	,1296	,1291	,1369	,1179	
24	,1707	,1639	,1664	,1662	,1662	,1610	,1706	,1756	,1744	,1708	,1246	,1249	,1251	,1136	,1364	,1364	,1469	,1240	
30	,1838	,1732	,1869	,1866	,1901	,1857	,1925	,1953	,1908	,1905	,1318	,1322	,1325	,1214	,1331	,1352	,1496	,1291	
36	,2045	,1776	,1974	,1976	,1977	,1911	,2038	,1997	,2107	,1996	,1517	,1508	,1511	,1259	,1367	,1391	,1560	,1319	

Tabla 5.9: Media de la raíz del error cuadrático medio para los 21 aerogeneradores de un parque eólico con una frecuencia de 10 minutos.

Valor relativo de la raíz del error cuadrático medio																
SETAR																
SETAR-TVCD																
h	ARMA	Media					Mediana					Media				
		TV-AR	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC
1	1,07	1,05	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2	1,17	1,08	1,13	1,15	1,18	1,14	1,15	1,26	1,28	1,18	1,01	1,13	1,18	1,00	1,06	1,07
3	1,22	1,13	1,25	1,25	1,25	1,22	1,29	1,25	1,34	1,22	1,12	1,12	1,12	1,00	1,06	1,07
4	1,26	1,17	1,26	1,26	1,26	1,24	1,32	1,36	1,31	1,30	1,08	1,08	1,08	1,00	1,14	1,13
5	1,30	1,21	1,30	1,30	1,30	1,30	1,33	1,41	1,39	1,39	1,07	1,07	1,07	1,00	1,17	1,16
6	1,32	1,25	1,34	1,35	1,35	1,36	1,39	1,40	1,42	1,39	1,09	1,09	1,09	1,00	1,17	1,16
12	1,42	1,31	1,38	1,36	1,39	1,36	1,43	1,41	1,44	1,41	1,10	1,10	1,10	1,00	1,17	1,16
18	1,54	1,37	1,41	1,41	1,42	1,40	1,48	1,48	1,50	1,47	1,10	1,10	1,10	1,00	1,17	1,16
24	1,62	1,40	1,47	1,46	1,46	1,42	1,50	1,55	1,54	1,50	1,10	1,10	1,10	1,00	1,20	1,20
30	1,68	1,43	1,54	1,54	1,57	1,53	1,59	1,61	1,57	1,57	1,09	1,09	1,09	1,00	1,10	1,11
36	1,76	1,41	1,57	1,57	1,57	1,52	1,62	1,59	1,67	1,58	1,20	1,20	1,20	1,00	1,09	1,10

Tabla 5.10: Valor relativo de la raíz del error cuadrático medio de 21 aerogeneradores de un parque eólico con una frecuencia de 10 minutos.

Media del error medio absoluto																
SETAR																
SETAR-TVCD																
h	ARMA	Media					Mediana					Media				
		TV-AR	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC
1	,0288	,0288	,0269	,0269	,0269	,0269	,0269	,0269	,0269	,0285	,0264	,0264	,0264	,0264	,0264	,0264
2	,0479	,0433	,0454	,0454	,0459	,0428	,0423	,0424	,0424	,0425	,0440	,0422	,0418	,0413	,0434	,0413
3	,0615	,0553	,0589	,0602	,0607	,0559	,0527	,0552	,0552	,0544	,0582	,0551	,0529	,0532	,0571	,0529
4	,0718	,0642	,0691	,0717	,0720	,0655	,0601	,0643	,0646	,0630	,0683	,0646	,0618	,0624	,0669	,0618
5	,0798	,0708	,0757	,0807	,0809	,0741	,0647	,0724	,0729	,0695	,0755	,0716	,0685	,0688	,0740	,0685
6	,0880	,0812	,0839	,0899	,0900	,0831	,0714	,0808	,0813	,0805	,0825	,0786	,0753	,0750	,0809	,0751
12	,1143	,0958	,1099	,1185	,1185	,1090	,0934	,1057	,1063	,0935	,1068	,1022	,0982	,0970	,1049	,0978
18	,1321	,1140	,1264	,1377	,1375	,1270	,1072	,1226	,1238	,1059	,1217	,1169	,1127	,1107	,1196	,1121
24	,1507	,1210	,1433	,1606	,1599	,1429	,1209	,1370	,1394	,1146	,1335	,1299	,1267	,1215	,1309	,1254
30	,1622	,1221	,1540	,1695	,1684	,1482	,1294	,1416	,1450	,1201	,1370	,1344	,1320	,1276	,1340	,1304
36	,1719	,1271	,1617	,1730	,1716	,1478	,1354	,1417	,1455	,1230	,1388	,1368	,1348	,1313	,1353	,1331

Tabla 5.11: Media del error absoluto medio para 21 aerogeneradores de un parque eólico con una frecuencia de 10 minutos.

		Valor relativo del error medio absoluto																				
		SETAR									SETAR-TVCD											
		Media						Mediana						Media						Mediana		
		TV-AR	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP	NT	NC	BETA	MEP
h	ARMA																					
1	1,09	1,09	1,02	1,02	1,02	1,02	1,02	1,02	1,02	1,08	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2	1,16	1,05	1,10	1,10	1,11	1,04	1,02	1,03	1,03	1,03	1,07	1,02	1,01	1,00	1,05	1,01	1,01	1,01	1,00	1,00	1,00	1,00
3	1,16	1,04	1,11	1,13	1,14	1,05	0,99	1,04	1,04	1,02	1,10	1,04	1,00	1,00	1,08	1,00	1,01	1,01	1,00	1,00	1,00	1,00
4	1,17	1,04	1,12	1,16	1,17	1,06	0,98	1,04	1,05	1,02	1,11	1,05	1,00	1,01	1,09	1,00	1,01	1,01	1,00	1,00	1,00	1,00
5	1,18	1,05	1,12	1,20	1,20	1,10	0,96	1,07	1,08	1,03	1,12	1,06	1,01	1,02	1,10	1,01	1,00	1,00	1,00	1,00	1,00	1,00
6	1,21	1,12	1,15	1,24	1,24	1,14	0,98	1,11	1,12	1,11	1,13	1,08	1,03	1,03	1,11	1,03	1,02	1,00	1,00	1,00	1,00	1,00
12	1,22	1,02	1,17	1,26	1,26	1,16	1,00	1,13	1,13	1,00	1,14	1,09	1,05	1,04	1,12	1,04	1,03	1,00	1,00	1,00	1,00	1,00
18	1,24	1,07	1,19	1,29	1,29	1,19	1,01	1,15	1,16	1,00	1,14	1,10	1,06	1,04	1,12	1,05	1,04	1,00	1,00	1,00	1,00	1,00
24	1,32	1,06	1,25	1,40	1,40	1,25	1,06	1,20	1,22	1,00	1,17	1,14	1,11	1,06	1,14	1,10	1,10	1,00	1,00	1,00	1,00	1,00
30	1,36	1,03	1,29	1,42	1,42	1,25	1,09	1,19	1,22	1,01	1,15	1,13	1,11	1,07	1,13	1,10	1,11	1,00	1,00	1,00	1,00	1,00
36	1,41	1,04	1,33	1,42	1,41	1,21	1,11	1,16	1,20	1,01	1,14	1,12	1,11	1,08	1,11	1,09	1,12	1,00	1,00	1,00	1,00	1,00

Tabla 5.12: Valor relativo del error medio absoluto para predicciones de 21 aerogeneradores de un parque eólico con una frecuencia de 10 minutos..

## Capítulo 6

# Contribuciones de la tesis

A lo largo de esta tesis se han ido mostrando resultados que nos han permitido cumplir el principal objetivo que nos planteamos al iniciarla, desarrollar técnicas para mejorar la implementación de modelos de series temporales no lineales, con especial énfasis en la predicción de potencia eólica.

El primer paso ha sido desarrollar una herramienta al estilo del correlograma en el ámbito de las series temporales lineales que nos ayude a identificar un modelo por umbrales. Para ello hemos propuesto una herramienta gráfica basada en la estimación recursiva del proceso ordenado respecto a la variable que provoca la no linealidad. Para comprobar si la estimación recursiva identificaba algún tipo de fenómeno no lineal hemos propuesto la construcción de una región de confianza dentro de la cuál cualquier ordenación de la variable que no provoque un cambio estructural en el proceso deberá de permanecer. Para la construcción de la región de confianza mostramos cuál sería la distribución asintótica del estimador para cualquier ordenación que se realice sobre el proceso.

Como extensión a la herramienta gráfica de identificación fue desarrollado un procedimiento que permite la modelización de estructuras TAR de manera automática. El procedimiento puede realizar la búsqueda de estructuras TAR empleando diferentes variables, tanto retardos del mismo proceso, como variables exógenas, como combinaciones de ellas.

Una vez desarrollado un procedimiento que nos permite la identificación de un tipo de modelo no lineal, estudiamos el problema de calcular predicciones empleando dichos modelos. Al hacerlo notamos que conseguir una buena estimación de la densidad predictiva era clave para obtener buenas predicciones, puesto que su cálculo implica la realización de simulaciones mediante un método de Monte Carlo. Por ello hemos propuesto una estimación adaptativa de la senda de densidades predictivas. Las principales novedades que introduce la estimación que proponemos son:

- (i) El uso de una estimación adaptativa de los momentos condicionados. En los trabajos previos sobre estimación de densidades predictivas, los momentos eran calculados empleando estimadores muestrales, sin tener en cuenta su posible evolución a lo largo del tiempo dependiendo del proceso.
- (ii) El uso de la distribución de Máxima Entropía. La distribución de Máxima Entropía ha sido empleada en el ámbito de la producción eólica para estimar la densidad incondicional de la velocidad del viento, pero no había sido empleada ni en datos de producción eólica ni en la estimación de densidades condicionales.

Por último, empleando el procedimiento automático de identificación de modelos TAR sobre datos de producción eólica encontramos un modelo SETAR con 6 regímenes. En este trabajo hemos mostrado cómo dicha estructura SETAR se encuentra como patrón común en datos de producción tanto horaria como diez minutal, así como en datos tanto de parques eólicos como de diferentes aerogeneradores de un mismo parque. Empleando la estructura común detectada hemos propuesto emplear para ajustar la producción eólica un modelo SETAR con distribución condicionada que va variando con el tiempo dependiendo del valor la potencia prevista. Para estimar dicha distribución condicionada introducimos el uso de la distribución de Máxima Entropía con heterocedasticidad condicional, en el cuál la varianza del proceso irá evolucionando dependiendo del valor de la potencia eólica prevista.

Las posibles líneas de investigación futuras a partir de los resultados aquí propuestos han sido mostradas a lo largo del trabajo, en las secciones [3.6](#), [4.8](#) y [5.5](#).







# Referencias

- [1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings of the Second International Symposium on Information Theory*. Akademiai Kiadó: Budapest, p. 267-281.
- [2] Akaike, H. (1977). On the entropy maximisation principle. In *Applications of Statistics* (ed. P.R. Krishnaiah). North-Holland: Amsterdam, p. 27-41.
- [3] Akpinar, S. y Kavak Akpinar, E. (2007). Wind Energy Analysis based on Maximum Entropy Principle (MEP)-type distribution function. *Energy Conversion and Management*, **48**: 1140-1149.
- [4] Al-Awami, A.T. y El-Sharkawi, M.A. (2009). Statistical Characterization of Wind Power Output for a given Wind Power Forecast. In *North American Power Symposium*, NAPS 2009.
- [5] Aldous, D. (1989). *Probability approximations via the Poisson clumping heuristic*. Springer-Verlag: New York.
- [6] Al-Qassam, M.S. y Lane, J.A. (1989). Forecasting exponential autoregressive models of order 1. *Journal of Time Series Analysis*, **10**: 95-113.
- [7] An, H.Z. y Huang, F.C. (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica*, **6**: 943-956.
- [8] Baïle, R., Muzy, J.F. y Poggi, P. (2011). Short-term forecasting of surface layer wind speed using a continuous random cascade model. *Wind Energy*, DOI: 10.1002/we.452.
- [9] Bellman, R.E. (1961). *Adaptive control processes*. Princeton University Press: Princeton.
- [10] Bermejo, M.A., Peña, D. y Sánchez, I. (2011). Identification of TAR models using recursive estimation. *Journal of Forecasting*, **30**: 31-50.

- [11] Bhattacharya, R.N. y Lee, C. (1995). Ergodicity of nonlinear first order autoregressive models. *Journal of Theoretical Probability*, **8**: 207-219.
- [12] Bludszuweit, H., Domínguez-Navarro, J.A. y Llombart, A. (2008). Statistical Analysis of Wind Power Forecast Error. *IEEE Transactions on Power Systems*, **23**: 983-991.
- [13] Bodewig, E. (1956). *Matrix Calculus*. North Holland: Amsterdam.
- [14] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**: 307-327.
- [15] Bossanyi, E.A. (1985). Short-term wind prediction using Kalman Filters. *Wind engineering*, **9**: 1-8.
- [16] Box, G. y Jenkins, G. (1970). *Time series analysis: Forecasting and control*, Holden-Day: San Francisco.
- [17] Bremnes, J.B. (2006). A Comparison of a Few Statistical Models for Making Quantile Wind Power Forecasts. *Wind Energy*, **9**: 3-11.
- [18] Brier, G.W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**: 1-3.
- [19] Brock, W.A., Scheinkman, J.A. y Dechert, W.D. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, **15**: 197-235.
- [20] Brown, B.G., Katz, R.W. y Murphy, A.H. (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of climate and applied meteorology*, **23**: 1184-1195.
- [21] Burton, T., Sharpe, D., Jenkins, N. y Bossanyi, E. (2001). *Wind Energy Handbook*. John Wiley & Sons: West Sussex.
- [22] Cai, Z., Fan, J. y Yao, Q. (2000). Functional-coefficient regression models for non-linear time series. *Journal of the American Statistical Association*, **95**: 941-956.
- [23] Campbell, E.P. (2004). Bayesian selection of threshold autoregressive models. *Journal of Time Series Analysis*, **25**: 467-482.
- [24] Carta, J.A., Ramirez, P. y Velázquez, S. (2009). A review of wind speed probability distributions used in wind energy analysis, Case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews*, **13**: 933-955.

- [25] Chan, K.S. (1986). *Topics in nonlinear time series analysis*. Doctoral thesis, Princeton University.
- [26] Chan, K.S. (1988). On the existence of the stationary and ergodic NEAR(p) model. *Journal of Time Series Analysis*, **9**: 319-328.
- [27] Chan, K.S. (1990). Testing for threshold autoregression. *The Annals of Statistics*, **18**: 1886-1894.
- [28] Chan, K.S. (1991). Percentage point of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society, Series B*, **53**: 691-696.
- [29] Chan, K.S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics*, **21**: 520-533.
- [30] Chan, K.S. y Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*. **7**: 179-190.
- [31] Chan, K.S. y Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society, Series B*, **52**: 469-476.
- [32] Chan, K.S. y Tsay, R.S. (1998). Limiting properties of the conditional least squares estimator of a continuous TAR model. *Biometrika*, **85**: 413-426.
- [33] Chan, K.S., Tong, H. & Wong, A.C.S. (2004). Some nonlinear threshold autoregressive time series models for actuarial use. *North American actuarial journal*, **8**: 37-61.
- [34] Chatfield, C. (2001). *Time Series Forecasting*, Chapman & Hall / CRC: London.
- [35] Chen, R. y Tsay, R.S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, **88**: 298-308.
- [36] Chen, R., Liu, J.S. y Tsay, R.S. (1995). Additivity tests for nonlinear autoregressions. *Biometrika*, **82**: 369-383.
- [37] Clements, M.P. y Harvey, D.I. (2010). Combining probability forecasts. *International Journal of Forecasting*, **27**: 208-223.
- [38] Clements, M.P. y Smith, J. (1997). The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting*, **13**: 463-475.
- [39] Clements, M.P., Franses, P.H., Smith, J. y Van Dijk, D. (2003). On SETAR non-linearity and forecasting. *Journal of Forecasting*, **22**: 359-375.

- [40] Cleveland, W.S. y McGill, R. (1987). Graphical Perception: The Visual Decoding of Quantitative Information on Statistical Graphs. *Journal of the Royal Statistical Society Series A*, **150**:192-229.
- [41] Cohen, A.C. (1957). On the solution of estimating equations for truncated and censored samples from normal populations. *Biometrika*, **44**: 225-236.
- [42] Cohen, A.C. (1991). *Truncated and Censored Samples*. Marcel Dekker: New York.
- [43] Corradi, V. y Swanson (2006). Predictive density evaluation. In *Handbook of Forecasting*, ed. por G. Elliot, C.W.J. Granger y A. Timmerman, Elsevier Science, 197-284.
- [44] Costa A. et al. 2008. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, **12**: 1725-1744.
- [45] Cox, D.R. (1961). Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society, Series B*, **23**: 414-422.
- [46] Cramer, H. (1961). On some classes of nonstationary stochastic processes. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **2**: 55-77.
- [47] Davies, R.B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **74**: 33-43.
- [48] Dawid, A.P. (1984). The prequential approach. *Journal of the Royal Statistical Society, Series A*, **147**: 278-292.
- [49] De Finetti, B. (1990). *Theory of Probability*. Wiley: New York.
- [50] De Gooijer, J.G. (2001). Cross-validation criteria for SETAR model selection. *Journal of Time Series Analysis*, **22**: 267-281.
- [51] De Gooijer, J.G. y De Bruin, P.T. (1998). On forecasting SETAR processes. *Statistics and Probability Letters*, **37**: 7-14.
- [52] De Gooijer, J.G. y Hyndman, R.J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, **22**: 443-473.
- [53] Diebold, F.X. y Lopez, J.A. (1996). Forecast evaluation and combination. In G.S. Maddala & C.R. Rao (Eds.), *Handbook of statistics*, **14**: 241-268. North Holland: Amsterdam.

- [54] Diebold, F.X., Gunther, T., Tay, A.S. (1998). Evaluating density forecasts with applications to finance and management. *International Economic Review*, **39**: 863-883.
- [55] Doukhan, P. (1994). *Mixing: Properties and Examples*. Springer-Verlag: New York.
- [56] Du, Z. y Escanciano, J.C. (2011). A simple portmanteau test for conditional goodness-of-fit in dynamic models. In *2011 North American Summer Meeting of the Econometric Society*, Washington University.
- [57] Epstein, E.S. (1969). Stochastic dynamic prediction. *Tellus*, **21**: 739-759.
- [58] Escanciano, J.C. (2006). Goodness-of-Fit Tests for Linear and Nonlinear Time Series Models. *Journal of the American Statistical Association*, **101**: 531-541.
- [59] Ezekiel, A. (1924). A method for handling curvilinear correlation for any number of variables. *Journal of the American Statistical Association*, **19**: 431-453.
- [60] Fan, J. y Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman and Hall: London.
- [61] Fan, J. y Yao, Q. (2005). *Nonlinear Time Series*. Springer: New York.
- [62] Fan, J., Zhang, C. y Zhang, J. (2001). Generalized likelihood test statistic and Wilks phenomenon. *The Annals of Statistics*, **29**: 153-193.
- [63] Fellows, A. y Hill, D. (1990). Wind and load forecasting for integration of wind power into a meso-scale electrical grid. *Proceedings of the European community Wind Energy conference*, 10-14.
- [64] Findley, D.F. (1983). On the use of multiple models for multi-period forecasting. In *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, p. 528-531.
- [65] Findley, D.F. (1985). Model selection for Multi-step-ahead forecasting. In *Identification and System Parameter Estimation (7th IFAC/IFORS Symposium)*, p 1039-1044, Pergamon Press: Oxford.
- [66] Fisher, R.A. (1935). *The Design of Experiment*. Hafner: New York.
- [67] Fortescue, T.R., Kershenbaum, L.S. y Ydstie, B.E. (1981). Implementation of self-tuning regulators with variable forgetting factors. *Automatica*, **17**: 831-835.
- [68] Franses, P.H. y van Dijk, D. (2000). *Non-linear time series models in empirical finances*. Cambridge University Press: Cambridge.

- [69] Fuller, W.A. y Hasza, D.P. (1981). Properties of predictors for autoregressive time series. *Journal of the American Statistical Association*, **76**: 155-161.
- [70] Galeano, P. y Peña, D. (2007). Improved model selection criteria for SETAR time series models. *Journal of Statistical Planning and Inference*, **137**: 2802-2814.
- [71] Gauss, C.F. (1809). *Theoria motus corporum celestium*. Traducido al inglés por C.H. Davis (1857). Reedición 1963, Dover Publications: New York.
- [72] Gaver, D.P. y Lewis, P.A.W. (1980). First order autoregressive gamma sequences and point processes. *Advances in Applied Probability*, **12**: 727-745.
- [73] Giebel, G., Brownsword, R. y Kariniotakis, G. (2011). The State-of-the-Art in Short-Term Prediction of Wind Power, Deliverable Report D1.1, Project Anemos, disponible en [http://anemos.cma.fr/download/ANEMOS\\_D1.1\\_StateOfTheArt\\_v1.1.pdf](http://anemos.cma.fr/download/ANEMOS_D1.1_StateOfTheArt_v1.1.pdf).
- [74] Giebel, G., Landberg, L., Nielsen, T.S. y Madsen, H. (2001). The Zephyr project. The next generation prediction system. *Proceedings of the EWEC*, 777-780.
- [75] Gneiting, T. y Raftery, A.E. (2007). Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association*, **102**: 359-378.
- [76] Gneiting, T., Balabdaoui, F. y Raftery, A.E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society, Series B*, **69**: 243-268.
- [77] González-Rivera, G., Senyuz, Z. y Yoldas, E. (2011). Autocontours: dynamic specification testing. *Journal of Business and economic statistics*, **29**: 186-200.
- [78] Granger, C.W.J. y Andersen, A.P. (1978). *An introduction to bilinear models*. Van derhoeck & Ruprecht: Gottingen.
- [79] Granger, C.W.J. y Teräsvirta, T. (1993). *Modeling nonlinear economic relationships*. Oxford University Press: Oxford.
- [80] Haggan, V. y Ozaki, T. (1981). Modeling nonlinear vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, **68**: 189-196.
- [81] Hamill, T.M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Reviews*, **129**: 550-560.
- [82] Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, **64**: 413-430.



- [83] Hansen, B.E. (1999). Threshold effects in non-dynamic panels: estimation, testing, and inference. *Journal of Econometrics*, **93**: 345–368.
- [84] Haykin, S., Sayed, A.H., Zeidler, J.R., Yee, P. y Wei, P.C. (1997). Adaptive tracking of linear time-variant systems by extended RLS algorithms. *IEEE Transactions on signal processing*, **45**: 1118-1128.
- [85] Hinich, M. (1982). Testing for gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*, **3**: 169-176.
- [86] Holt, C.C. (1957). *Forecasting trends and seasonals by exponentially weighted moving averages*, O.N.R. Memorandum, n°52, Carnegie Institute of Technology.
- [87] Hong, Y. (1997). One-sided testing for conditional heteroskedasticity in time series models. *Journal of Time Series Analysis*, **18**: 253-277.
- [88] Hong-zhi, A. y Bing, C. (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *International Statistical Review*, **59**: 287-307.
- [89] Hurvich, C.M. y Tsai, C.L. (1989). Regression and time series model selection in small samples. *Biometrika*, **76**: 297-307.
- [90] Hurvich, C.M., Shumway, R. y Tsai, C.L. (1990). Improved estimators of Kullback-Leibler information for autoregressive model selection in small samples. *Biometrika*, **77**: 709-719.
- [91] Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical Reviews*, **106**: 620-630.
- [92] Jolliffe, I.T. y Stephenson, D.B. (2003). *Forecast verification: a practitioner's guide in atmospheric science*. Wiley: Chichester.
- [93] Jones, D.A. (1976). *Non-linear autoregressive processes*. PhD Thesis, University of London.
- [94] Kamal, L. y Jafri, Y.Z. (1997). Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. *Solar Energy*, **61**: 23-32.
- [95] Kapetanios, G. (2001). Model selection in threshold models. *Journal of Time Series Analysis*, **22**: 733-754.

- [96] Kariniotakis, G., Nogaret, E. y Stavrakis, G. (1997). Advanced short-term forecasting of wind power production. *Proceedings of the European Wind Energy conference*, 751-754.
- [97] Kariniotakis, G., Nogaret, E. y Stavrakis, G. (1999). Evaluation of advanced Wind Power and Load forecasting methods for the Optimal Management of Isolated Power Systems. *Proceedings of the European Wind Energy conference*, 1082-1085.
- [98] Keenan, D.M. (1985). A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika*, **72**: 39-44.
- [99] Kunitomo, N. y Yamamoto, T. (1985). Properties of predictors in misspecified autoregressive time series models. *Journal of the American Statistical Association*, **80**: 941-950.
- [100] Landau, I.D., Lozano, R. y M'Saad, M. (1998). *Adaptive control*, Springer: London.
- [101] Landberg, L., Giebel, G., Madsen, H., Nielsen, T.S., Jorgensen, J.U., Laursen, L. et al. (2002). *Wind farm production prediction -the Zephyr model*. Technical report. Roskilde, Denmark: Riso national Laboratory.
- [102] Lange, M. (2005). On the Uncertainty of Wind Power Predictions: Analysis of the Forecast Accuracy and Statistical Distribution of Errors. *Journal of Solar Energy Engineering*, **127**: 177-184.
- [103] Lau, A. y McSharry, P. (2010). Approaches for Multi-Step Density Forecasts with Application to Aggregated Wind Power. *The Annals of Applied Statistics*, **4**: 1311-1341.
- [104] Lawrance, A.J. y Lewis, P.A.W. (1980). The Exponential Autoregressive-Moving Average EARMA (p,q) Process. *Journal of the Royal Statistical Society. Series B (Methodological)*, **42**: 150-161.
- [105] Lee, J.H.H. y King, M.L. (1993). A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business and Economic Statistics*, **11**: 17-27.
- [106] Legendre, A.M. (1805). Nouvelles méthodes pour la détermination des orbites des comètes, en *Sur la méthode des moindres carrés*. París.
- [107] Li, W.K. y Lam, K. (1995). Modelling asymmetry in stock returns by a threshold ARCH model. *The Statistician*, **44**: 333-341.

- 
- [108] Ling, S. y Tong, H. (2005). Testing for a linear MA model against threshold MA models. *Annals of Statistics*, **33**: 2529-2552.
- [109] Ling, S. y Tong, H. (2011). Score based goodness-of-fit tests for time series. *Statistical Sinica*, Preprint SS-09-090.
- [110] Liu, F.J. y Chang, T.P. (2010). Validity analysis of maximum entropy distribution based on different moment constraint for wind energy assessment. *Energy*, doi:10.1016/j.energy.2010.11.033.
- [111] Ljung, L. y Söderström, T. (1983). *Theory and practice of recursive identification*. The MIT Press: Cambridge.
- [112] McLeod, A.I. y Li, W.K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, **4**: 269-273.
- [113] Maravall, A. (1983). An application of nonlinear time series forecasting. *Journal of the Business and Economic Statistics*, **1**: 66-74.
- [114] Matheson, J.E. y Winkler, R.L. (1976). Scoring rules for continuous probability distributions. *Management science*, **22**: 1087-1096.
- [115] McQuarrie, A., Shumway, R. y Tsai, C.L. (1997). The model selection criterion AICu. *Statistics and Probability Letters*, **34**: 285-292.
- [116] Mitchell, J. (2008). Density forecast revisions and forecast efficiency. *National Institute of Economic and Social Research*.
- [117] Moller, J.K., Nielsen, H.A. y Madsen, H. (2008). Time-adaptive quantile regression. *Computational Statistics & Data Analysis*, **52**: 1292-1303.
- [118] Moran, P.A.P. (1953). The statistical analysis of the Canadian lynx cycle, I: structure and prediction. *Australian Journal of Zoology*, **1**: 163-173.
- [119] Nielsen, T.S., Madsen, H. y Tofting, J. (1999). Experiences with statistical methods for wind power prediction. *Proceedings of the EWEC*, 1066-1069.
- [120] Nielsen, H.A., Madsen, H., Nielsen, T.S., Badger, J., Giebel, G., Landberg, L., Sattler, K. y Feddersen, H. (2004). *Wind Power Ensemble Forecasting*. Technical report, Technical University of Denmark, DTU Informatics, Kgs. Lyngby, Denmark, 2004.
- [121] Öhrvik, J. y Schoier, G. (2005). SETAR model selection, a bootstrap approach. *Computational Statistics*, **20**: 559-573.

- [122] Ozaki, T. y Oda, H. (1978). Non-linear time series model identification by Akaike's information criterion. *Proceedings IFAC Workshop on Information and Systems*, Compiegne, France.
- [123] Pankratz, A. (1983). *Forecasting with univariate Box-Jenkins models: concepts and cases*. Wiley: Hoboken.
- [124] Peña, D. (2002). *Análisis de datos multivariantes*, McGraw-Hill: Madrid.
- [125] Peña, D. y Rodríguez, J. (2005). The Log of the determinant of the autocorrelation matrix for testing goodness of fit in time series. *Journal of Statistical Planning and Inference*, **136**: 2706-2718.
- [126] Peña, D. y Rodríguez, J. (2005b). Detecting nonlinearity in time series by model selection criteria. *International Journal of Forecasting*, **21**: 731-748
- [127] Peña, D., Tiao, G.C. y Tsay, R.S. (2001). *A course in time series analysis*. Wiley: New York.
- [128] Petrucci, J. y Davies, N. (1986). A portmanteau test for Self-Exciting Threshold Autoregressive-Type nonlinearity in time series. *Biometrika*, **73**: 687-694.
- [129] Pinson, P. (2010). *On probabilistic Forecasting of Wind Power Time-Series*. Technical report, Technical University of Denmark, DTU Informatics, Kgs. Lyngby, Denmark, 2010.
- [130] Pinson, P. y Madsen, H. (2009). Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy*, **12**: 137-155.
- [131] Pinson, P., Christensen, L.E.A., Madsen, H., Sorensen, P.E., Donovan, M.H. y Jensen, L.E. (2008). Regime-switching modelling of the fluctuations of offshore wind generation. *Journal of Wind Engineering and Industrial Aerodynamics*, **96**: 2327-2347.
- [132] Pitman, E.J.G. (1938). Significance tests which may be applied to samples from any population. *Biometrika*, **29**: 322-335.
- [133] Potter, S. (1995). A nonlinear approach to US GNP. *Journal of Applied Econometrics*, **10**: 109-125.
- [134] Priestley, M.B. (1980). State dependent models: a general approach to nonlinear time series analysis. *Journal of Time Series Analysis*, **1**: 57-71.
- [135] Priestley, M.B. (1981). *Spectral Analysis and Time Series*. Academic Press: New York.

- [136] Ramírez, P. y Carta, J.A. (2006). The use of Wind Probability Distributions derived from the Maximum Entropy Principle in the Analysis of Wind Energy. A case study. *Energy Conversion and Management*, **47**: 2564-2577.
- [137] Ramsey, J.B. (1969). Tests for Specification Errors in Classical Linear Least Squares Regression Analysis. *Journal of the Royal Statistical Society Series B*, **31**: 350-371.
- [138] Rao Sripada, N. y Gran Fisher, D. (1987). Improved least squares identification. *International Journal of Control*, **46**: 1889-1913.
- [139] Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The annals of mathematical statistics*, **23**: 470-472.
- [140] Samia, N.I., Chan, K.S. y Stenseth, N.C. (2007). A generalized threshold mixed model for analyzing nonnormal nonlinear time series; with application to plague in Kazakhstan. *Biometrika*, **94**: 101-118.
- [141] Sánchez, I. (2006). Recursive estimation of dynamic models using Cook's distance, with application to wind energy forecast. *Technometrics*, **48**: 61-73.
- [142] Sánchez, I. (2006b). Short-term prediction of Wind Energy production. *International Journal of Forecasting*, **22**: 43-56.
- [143] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**: 461-464.
- [144] Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**: 379-423, 623-656.
- [145] Shorack, G.R. y Wellner, J.A. (1986). *Empirical processes with applications to statistics*. Wiley: New York.
- [146] Siddall, J.N. y Diab, Y. (1975). The use in probabilistic design of probability curves generated by maximizing the Shannon entropy function constrained by moments. *Journal of Industrial Engineering*, **97**: 843-852.
- [147] Stenseth, N.C., Chan, K.S., Tong, H. et. al. (1999). Common dynamic structure of Canadian lynx populations within three climatic regions. *Science*, **285**: 1071-1073.
- [148] Stenseth, N.C., Samia, N.I. et. al. (2006). Plague dynamics are driven by climate variation. *Proceedings of the National Academy of Sciences*, **103**: 13110-13115.
- [149] Stoica, P., Eykhoff, P., Janssen, P. y Söderström, T. (1986). Model-structure selection by cross-validation. *International Journal of Control*, **43**: 1841-1878.

- [150] Stone, M. (1974). Cross-validatory choice and assessment of statistical prediction. *Journal of Royal Statistical Society, Series B (Methodological)*, **36**: 111-147.
- [151] Subba Rao, T. y Gabr, M.M. (1980). A test for linearity of stationary time series. *Journal of Time Series Analysis*, **1**: 145-158.
- [152] Subba Rao, T. y Gabr, M.M. (1984). *An introduction to bispectral analysis and bilinear time series models*. Springer-Verlag: New York.
- [153] Tantareanu, C. (1992). *Wind prediction in Short term: A first stem for a better wind turbine control*. Nordvestjysk Folkecenter for Vedvarende Energi.
- [154] Tay, A.S. y Wallis, K.F. (2000). Density Forecasting: A Survey. *Journal of Forecasting*, **19**: 235-254.
- [155] Taylor, S.J. (1987). Forecasting the volatility of currency exchange rates. *International Journal of Forecasting*, **3**: 159-170.
- [156] Taylor, J.W., McSharry, P.E. y Buizza, R. (2009). Wind Power Density Forecasting Using Wind Ensemble Predictions and Time Series Models. *IEEE Transactions on Energy Conversion*, **24**: 775-782.
- [157] Tiao, G.C. y Tsay, R.S. (1994). Some advances in nonlinear and adaptive modeling in time series. *Journal of Forecasting*, **13**: 109-131.
- [158] Tjøstheim, D. (1986). Some doubly stochastic time series models. *Journal of Time Series Analysis*, **7**: 51-72.
- [159] Tong, H. (1977). Discussion of a paper by A.J. Lawrance and N.T. Kottegoda. *Journal of the Royal Statistical Society, Series B*, **140**: 34-35.
- [160] Tong, H. (1983). *Threshold models in non-linear time series analysis*. Springer: New York.
- [161] Tong, H. (1987). Non-linear time series models of regularly sampled data: A survey. *Proceedings of the First World Congress of the Bernoulli Society, 1986*. Editado por Y.V. Prohrov y V.V. Sansonov, **2**, p.355-367. VNU Science Press: Amsterdam.
- [162] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press: Oxford.
- [163] Tong, H. (2010). *Threshold models in time series analysis - 30 years on*. Research Report. Serial N° 471. The University of Hong Kong.

- 
- [164] Tong, H. y Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data (con discusión). *Journal of Royal Statistical Society, Series B (Methodological)*, **42**: 245-292.
- [165] Tong, H., Thanoon, B. y Gudmundsson, G. (1985). Threshold time series modelling of two Icelandic riverflow systems. *Water Resources Bulletin*, **21**: 651-661.
- [166] Torres, J.L., Garcia, A., De Blas, M. y De Francisco, A. (2005). Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Solar Energy*, **79**: 65-77.
- [167] Tsay, R.S. (1986). Nonlinearity tests for time series. *Biometrika*, **73**: 461-466.
- [168] Tsay, R.S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, **84**: 231-240.
- [169] Tukey, J.W. (1949). One degree of freedom for non-additivity. *Biometrika*, **5**: 232-242.
- [170] Tyssedal, J.S. y Tjøstheim, D. (1988). An autoregressive model with suddenly changing parameters and an application to stock market prices. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **37**: 353-369.
- [171] Unnikrishnan, N.K. (2004). Bayesian subset model selection for time series. *Journal of Time Series Analysis*, **25**: 671-690.
- [172] Van der Leeuw, J. (1994). The covariance matrix of ARMA errors in closed form. *Journal of Econometrics*, **63**: 397-405.
- [173] Volterra, V. (1930). *Theory of functional and of integro-differential equations*. Dover: New York.
- [174] Wedlin, A. (1997). On the notion of second-order exchangeability. *Erkenntnis*, **45**: 177-194.
- [175] Whittle, P. (1954). The statistical analysis of a seiche record. *Sears Foundation Journal of Marine Research*, **13**: 76-100.
- [176] Wong, C.S. y Li, W.K. (1998). A note on the corrected Akaike information criterion for the threshold autoregressive models. *Journal of Time Series Analysis*, **19**: 113-124.
- [177] Wu, S. y Chen, R. (2007). Threshold variable determination and threshold variable driven switching autoregressive models. *Statistica Sinica*, **17**: 241-264.